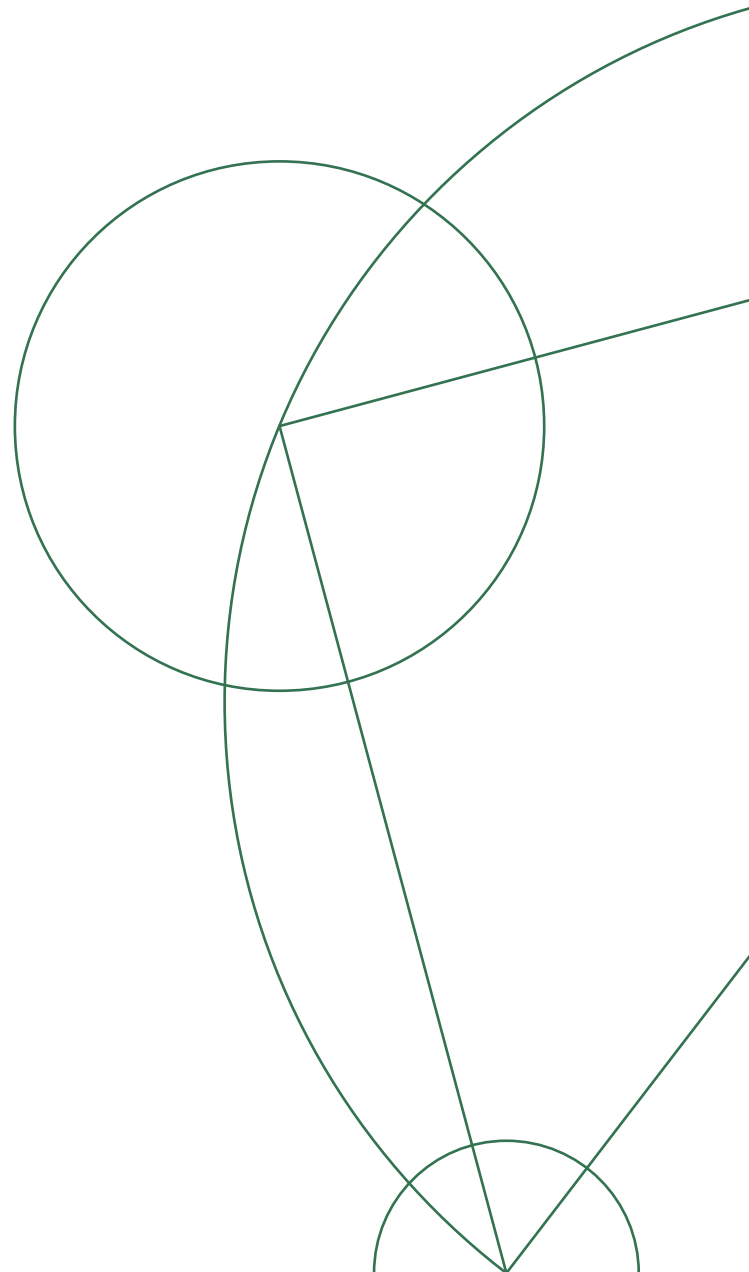FACULTY OF SCIENCE
UNIVERSITY OF COPENHAGEN

Electrostatic simulations of nanoelectronic hybrid devices

Master's Thesis in Physics
Waldemar Svejstrup
May 20, 2021

Supervised by
Karsten Flensberg

Abstract

In this thesis, we have studied electrostatic simulations of nanoelectronic semiconductor-superconductor hybrid devices. We have numerically done this using the widespread Thomas-Fermi and Schrödinger-Poisson methods. On top of this, we have constructed an orbital-free approach similar to the Thomas-Fermi method but including the von Weizsäcker correction term to the energy functional. This correction takes the spatial modulation of the electron density into account and thus extends the local Thomas-Fermi method with the non-local von Weizsäcker term, forming the new extended Thomas-Fermi method. We found that its quality of results and computational speed lie between the Thomas-Fermi and Schrödinger-Poisson methods, allowing for more compromise between accuracy and computational cost. We compare these three numerical methods and apply them to simulate the electrostatics of real hybrid devices.

Acknowledgements

Contents

Chapter 1

Introduction

Over the years, computers have become faster and faster, and their components smaller and smaller [1]. This evolution has led us to a point at which the single components of the modern day computer are so small, that quantum effects start to play a role. This is predicted by Moore's law [2], which states that the pattern size in computer chips will follow an exponential decrease over time. Since the first transistor was invented in 1949, their physical size has been exponentially decreasing as predicted by Moore, which means that we are now pushing the physical transistor size to a limit that invokes quantum physics [2]. This can indeed be viewed as both an advantage and a disadvantage. The disadvantage being that the classical way of making computer bits will face a constraint on the physical size, since the quantum effects will dominate at a certain component size. As this gives a lower bound for how small each component can be made, and as there are limits as to how many components can actually be integrated on a chip, the classical computer will at some point reach a state where it simply cannot become much faster. The advantage is to use these quantum effects to construct a new type of bit, the so-called quantum bit (qubit), which we will see can be superior to the classical bit in some aspects [1, 3].

1.1 Nanoelectronic hybrid devices and quantum computation

Nanoelectronic hybrid devices use the physics induced by the proximity effect in the close vicinity of superconductor-semiconductor interfaces. Exploiting exactly this hybridization allows for conventional s-wave superconductivity in the semiconductor [4]. It has been theoretically predicted that the induced superconductivity can be combined with spin-orbit coupling and an applied magnetic field to drive the system into a p-wave superconducting state [4]. This p-wave superconducting state is topologically non-trivial and thus of great scientific interest, as it supports charge-neutral zero-energy states obeying non-Abelian exchange statistics [4]. These states are called Majorana zero modes (MZMs) and form a fascinating foundation for topological quantum computing. The details of the hybrid devices themselves and topological quantum computation are discussed below. Before starting this exciting discussion of hybrid devices and their applications, we will briefly discuss quantum computation and its relevance.

One application of quantum computation, as shown by Feynmann in 1982 [1], is that one can simulate certain quantum systems much more effectively on a quantum computer, as certain many-body quantum Hamiltonians are simulated exponentially faster on a quantum computer [1].

Another significant application, as discovered by Peter Shor in 1994 [1, 5], is a

quantum algorithm than can find the prime factor of an m digit number in a polynomial timescale. This can be much faster than the exponential time it takes the fastest algorithm for classical computers [1]. This is very relevant for encryption and has caused much interest, not only in the physics community.

Other applications are, for instance, Grover's algorithm, which through quantum computation gives a great speed up in database search compared to search algorithms on a classical computer [6]. The quantum computer brings a vast number of applications and opportunities. It is beyond the scope of this thesis to discuss the details of the possible applications. Instead, we will start our discussion at the heart of a quantum computer, the *qubit*, and investigate how to realise such qubits. Furthermore, we will investigate and develop numerical methods that can be useful for constructing topologically protected qubits.

Classical computers consist of classical bits. Every bit is a simple classical two-level system that can be either "0" or "1" [1]. Such a two-level system can be created quantum mechanically as well. However, a quantum two-level system is no longer required to be either "0" or "1". Actually, a quantum two-level system can be in one of infinitely many superpositions, i.e. $a\left|0\right\rangle + b\left|1\right\rangle$ [1]. This means that for n qubits, the state will be a vector in a 2^n dimensional Hilbert space, where all the qubits can be entangled with each other. This superposition allows a system to traverse many trajectories in parallel and thus determine the final state [1]. This is an extraordinary property of qubits that gives many advantages over classical bits.

The requirements to realise quantum information systems can be confined to two general characteristics [2]:

- A Hilbert space must be spanned by orthonormal quantum states, such as the $\left|0\right\rangle$ and $\left|1\right\rangle$ states just introduced. This can, of course, be generalized to arbitrarily many orthonormal states $\left|n\right\rangle$. A density matrix can then describe the state of any particular system in this Hilbert space [2].

- There must exist a probability distribution, call it p_n, that allows us to write the state of the system as $\hat{p} = \sum_n p_n \left|n\right\rangle \left\langle n\right|$ [2].

However, as we shall see, there are several obstacles in realising such a system of entangled qubits. Two of the obstacles we will briefly describe are handling random and systematic errors in quantum computers. One should note that error-correcting schemes are possible for quantum computers [1], nevertheless, random and systematic errors are still a far-reaching issue. Random errors are generally caused by interactions with the environment [1]. This causes the quantum computer, initially in a pure superposition state, to become entangled with the environment. This means that the environment has caused decoherence of the quantum computer's state, and quantum information is lost [1]. Systematic errors are unitary errors occurring while processing quantum information. For instance, one could aspire to a 90-degree qubit rotation but might unknowingly end up rotating by 90.01 degrees [1]. However, if qubits were topologically protected, they would be robust to these kinds of errors [1]. This is just one of the useful properties that can be exploited in hybrid devices, as discussed in the sections below.

1.2   Topological quantum computation

We will start our discussion of topologically protected qubits with a brief introduction to Majorana fermions and MZMs. A Majorana fermion is a fermionic particle that is its own antiparticle [7]. A MZM is a Majorana fermion bound to a defect at 0 energy. MZMs are non-Abelian particles, which allow them to be the foundation of topologically protected quantum computations [7]. In contrast to bosons, which have several particles that are their own antiparticles, it is quite a unique characteristic for a fermion to be its own antiparticle. Mathematically, this means that the Majorana creation operator γ_1 is a Hermitean operator $\gamma_1 = \gamma_1^\dagger$, that it anticommutes with other Majorana creation operators γ_2, and that it squares to one [7]. This means that the Majorana operators form a Clifford algebra, which is defined by the anticommutation relation $\{\gamma_k, \gamma_l\} = 2\delta_{kl}$ [7]. One challenge is that Majorana fermions do not exist as elementary particles, and thus we need to construct them from conventional Dirac fermions created by $c_k^\dagger$, with canonical anticommutation relations $\{c_k, c_l^\dagger\} = \delta_{kl}$ and $\{c_k, c_l\} = 0$ [7]. It can be shown that for $k = 1, ..., N$ Dirac fermions, we can construct $2N$ Majorana operators such that [7]

$$\gamma_{2k-1} = c_k + c_k^\dagger \quad \text{and} \quad \gamma_{2k} = i(c_k^\dagger - c_k). \tag{1.1}$$

The Hilbert space of a single fermionic particle is two-dimensional, as it is either empty or filled. This is distinguished by the eigenvalue of the number operator $n_k = c_k^\dagger c_k$ [7], which can be shown to be either 0 or 1. From the number operator, one can construct the so-called parity operator $\mathcal{P}_k = 1 - 2n_k = (-1)^{n_k} = -i\gamma_{2k-1}\gamma_{2k}$, which has eigenvalue 1 if the number of fermionic modes is even, and -1 if it is odd [7]. So far the introduction of the Majorana operators has just been an algebraic way of rewriting the complex operators c_k in terms of Hermitian operators γ_k. However, it turns out that these operators actually describe the excitations of a certain physical Hamiltonian [7, 8].

Considering the parity operator $\mathcal{P}_k$ again, this is the parity operator for the pair of MZMs γ_{2k-1} and γ_{2k}. One can also define the total fermion parity $\mathcal{P}_{\text{tot}} = \prod_k \mathcal{P}_k$. The total fermionic parity operator has eigenvalues ± 1, depending on whether the total number of occupied fermionic modes is even or odd [9]. If we consider a system of N pairs of MZMs described by the states $|\Psi\rangle$, it is only meaningful to consider the states $|\Psi\rangle$ that are eigenstates of $\mathcal{P}_{\text{tot}}$, i.e. $\mathcal{P}_{\text{tot}}|\Psi\rangle = \pm|\Psi\rangle$ [9]. This means that all linear combinations of states with a different total parity are forbidden for a closed system. Assuming now that we can exchange any two MZMs in this system consisting of N pairs of MZMs. We can then swap the position of two MZMs, leaving the final configuration identical to the initial configuration. If the swap takes time T, we must require that the Hamiltonian governing the swap obeys $H(0) = H(T)$ [9]. To ensure that the configuration does not leave the initial ground state, we must also require that we change the Hamiltonian slowly enough to obey the adiabatic theorem [9]. This corresponds to some unitary operator U, such that $|\Psi\rangle \rightarrow U|\Psi\rangle$ [9]. We should again note, that the adiabatic exchange of two MZMs does not change the total parity, meaning mathematically that $[U, \mathcal{P}_{\text{tot}}] = 0$. It can further be shown, that the unitary operator that exchanges two MZMs takes the form $U = \frac{1}{\sqrt{2}}(1 \pm \gamma_n \gamma_m)$, where $\pm$ distinguishes

between a clockwise and a counterclockwise exchange [9].

To see the effect of this, consider 4 MZMs γ_1, γ_2, γ_3, and γ_4. Their groundstate manifold has four different states, $|00\rangle$, $|01\rangle$, $|10\rangle$ and $|11\rangle$, where the first digit is the occupation number of the fermionic mode $c_1^\dagger = \frac{1}{2}(\gamma_1 + i\gamma_2)$ and the second digit is the occupation number of the fermionic mode $c_2^\dagger = \frac{1}{2}(\gamma_3 + i\gamma_4)$ [9]. If we start in the state $|00\rangle$, and exchange MZM 2 and 3 (choosing now the clockwise direction such that we get $+$ for simplicity), it can be shown that $U_{23}|00\rangle = \frac{1}{\sqrt{2}}(|00\rangle + i|11\rangle)$ [9]. This means that the exchange of two MZMs results in a superposition of states, and not just an overall phase as we see for bosons and fermions [9]. Furthermore, it can be shown that if a sequence of exchanges is made, the ordering matters. If we, for instance, consider the four MZMs again, it is not the same to exchange MZM 2 and 3, and afterwards 1 and 2, as to do the opposite ordering, $U_{23}U_{12} \neq U_{12}U_{23}$ [9].

To summarise, we see that the exchange of two MZMs gives a non-trivial rotation in the ground-state manifold and that the ordering of these exchanges matters [9]. These properties make MZMs *non-abelian anyons*. The exchange of any two MZMs is often called *braiding*, as the trajectories of sequential exchanges look like a braid, when plotted as a worldline.

These non-abelian properties are indeed very special properties. Regarding quantum computation, the non-abelian properties allow different sequences of MZM exchange operations to represent different algorithms [9]. The perk of doing quantum algorithms this way lies in the topological protection. The state of the system is encoded in the fermion parity degrees of freedom, which are shared non-locally among the MZMs [9]. This means that no local perturbation can cause decoherence of the quantum state, and thus the quantum state is topologically protected by the parity [9]. As we execute the algorithms by exchanging the MZMs, we also get around the error described above where one might unknowingly rotate a state for instance 90.01 degrees instead of 90 degrees. We get around this because there is no such thing as almost exchanging two MZMs, and thus every braid operation corresponds to some *exact* rotation of the groundstate manifold [9].

With these arguments, it should be clear that it is very sought-after to realise systems in which MZMs exist. This thesis deals with nanoelectronic hybrid devices, which are theorised to serve as a platform for MZMs. Before discussing hybrid devices in greater detail, we will briefly discuss some of the superconducting characteristics needed to realise Majorana modes in hybrid devices.

1.2.1 Topology and Andreev reflection

Firstly, we will consider the transition between a trivial and a topological superconducting phase. This can be understood from studying the band structure of the superconductor. On an abstract level, a topological system will be described by some Hamiltonian, which leads to energy levels with some energy gap [9]. Another Hamiltonian will then be topologically equivalent if it also has some energy gap, and the two

Hamiltonians can be deformed into each other by a continuous transformation [9]. The topic of topological superconductivity is indeed fascinating and far-reaching. For now, we will settle with the knowledge that we can drive a system from a topological phase to a trivial phase, and the other way around, by adjusting the system parameters [9].

We will now consider an interface with a normal metal and a superconductor (so-called NS interface). If we apply a voltage over the system, we would naturally expect a current to run. However, if the voltage applied (V) obeys that $eV < \Delta$, where e is the electron charge, and Δ is the superconducting gap, it may not be so trivial what happens [9]. The voltage applied is not enough to fill states in the superconductor, as it would require energy larger than the superconducting gap, and thus one might be lead to the belief that no current can run. In the regime where $eV < \Delta$, it is actually possible for us to get a current through the system, which happens by Andreev Reflection [9]. Here an incident electron coming from the normal metal is converted to a hole (travelling opposite of the incident electron), and a Cooper pair is created in the superconductor [9]. This is schematically shown in Fig. 1.1.
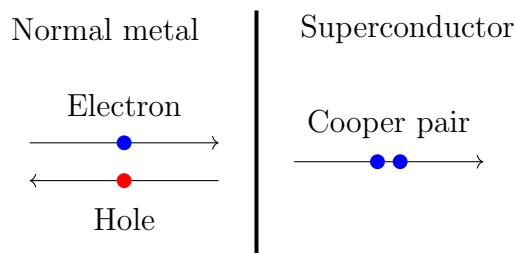
Figure 1.1: Illustration of Andreev reflection. An incident electron (blue dot) coming from the left is converted to a hole (red dot) and a Cooper pair (double blue dots) [9].

Because both the hole and the electron in the normal lead contribute to the transfer of charge, this can conceptually be considered as a transmission problem. It can be shown that the conductance $G(V) = dI/dV$ is given by [9]

$$G(V) = 2G_0|r|^2, \tag{1.2}$$

where $G_0 = e^2/h$ is the conductance quantum, the factor of 2 is from the transfer of Cooper pairs, and $|r|^2$ is the probability of Andreev Reflection.

This becomes very relevant to our specific use if the superconductor is in a topological phase and there is a MZM on the edge of the superconductor at the NS interface. Considering the transmission case from the Andreev reflection, it can be shown that a resonant transmission is obtained in the presence of a bound state [9]. Thus the probability $|r|^2$ is greatly enhanced if the energy of the incident electron matches the one of the MZM. As the energy of the MZM is zero, we will expect a resonant peak in conductance of the NS interface at $V = 0$ [9]. In the case of perfect Andreev reflection $|r|^2 = 1$, we will thus have a peak in conductance of $G = 2G_0$. This is a so-called zero-bias peak and is generally believed to be one of the strongest signs of MZMs [9]. This is further described in the discussion of hybrid devices.

1.3 Details of hybrid devices

The hybrid devices considered in this thesis consist of a semiconducting stack with a superconducting wire on top. On top of the semiconducting stack and the superconducting wire, oxide is grown via atomic layer deposition, and on top of the oxide, gates are defined via electron beam lithography. These gates make it possible to control the electrostatics of the device, such that the device becomes a suitable platform to search for MZMs.

The gates in the hybrid devices considered in this thesis are all Ti/Au gates isolated from the semiconducting stack and the superconducting wire by a layer of dielectric oxide. The superconducting wires are all made from aluminium and in direct contact with the semiconducting stack. These wires are considered Ohmic gates, i.e. they form an Ohmic contact when in contact with the semiconducting stack [2, 10]. This means that there is no interfacial barrier on the edge of the wire, in contrast to the widely used Schottky-contact, which the so-called Schottky-barrier characterizes. A Schottky contact will have a barrier limiting the transport of electrons (or holes), which is physically caused by an accumulation of surface states on the wire [2]. At the interface between aluminium and semiconductor in the hybrid devices, we will assume that the conduction band minimum of the semiconductor is below the Fermi energy of the aluminium. This means that electrons from the aluminium will move into the semiconductor, thus causing band bending. This is further discussed in section 2.4.1. Another thing to note about the hybrid devices considered in this thesis is that all the semiconductors are considered intrinsic. Thus they are neither p-doped nor n-doped. A simplified cross-section of a typical device can be seen in Fig. 1.2.
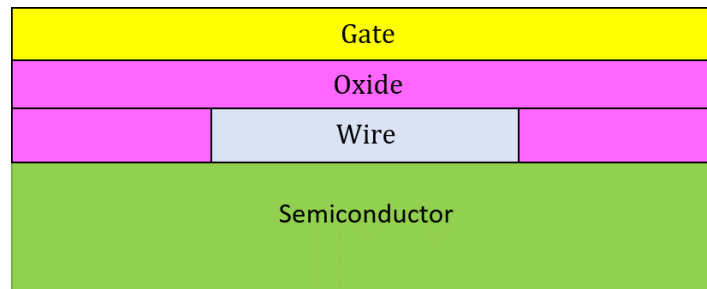
Figure 1.2: Illustration of a typical cross section of the hybrid devices investigated in this thesis. Note how the wire (grey) is isolated from the gate (yellow). The semiconducting stack (green) is a heterostructure that forms a quantum well.

The electrostatics controlled by the gates confine the electrons in the semiconducting stack to the area below the wire, i.e. the gates create lateral confinement by depleting all electrons not directly below the wire. This happens as negative voltage is applied to the gates, driving the electrons away from the gates. The wire screens the effect of the gates, and thus the electron density remains only underneath the wire. Another confinement happens in the semiconducting stack, as this consists of several different semiconductors that together form a quantum well. The confinement caused by the quantum well forms a 2-dimensional electron gas (2DEG), and the gates thus confine

the 2DEG to the area below the superconducting wire, effectively creating a quasi-1D electron wire with induced superconductivity. This confinement is further discussed in section 2.4.1. Quantum point contacts (QPCs) and quantum dots can then be formed by applying suitably shaped gates that affect the confined 2DEG.

The superconducting wire induces superconductivity in the semiconductor beneath it via the proximity effect [11]. The proximity effect is caused by Andreev reflection of quasiparticles at the interface, forming correlated electron-hole pairs and thus inducing superconductivity in the semiconductor [11]. Applying a Zeeman field to the system, we expect to induce a topological superconducting gap accompanied by a pair of localized MZMs at each end of the 2DEG wire [12, 13]. This happens as the magnetic field can be increased in order to drive the system through a topological quantum phase transition, taking the system from a trivial to a topological superconducting phase [13]. We expect the bulk quasiparticle gap to vanish and reopen at the topological quantum phase transition as a topological gap [13]. It is this topological gap that hosts the MZMs.

In order to detect MZMs, one can measure the charge tunnelling current into the edge of the 2DEG, which is supposed to result in a zero-bias conductance peak (ZBCP) [12, 13, 14]. Thus a signature of MZMs (but not proof [15]) could be a reopening of the bulk gap at the topological quantum phase transition simultaneous with the appearance of a ZBCP [12]. One would expect this sharp transition between trivial and topological phases to take place at $E_z = \sqrt{\Delta_{\mathrm{ind}}^2 + \mu^2}$, where Δ_{ind} is the superconducting gap in the wire, μ is the chemical potential, and E_z is the Zeeman splitting given by $E_Z = g\mu_B B$ where g is the Landé g factor, μ_B is the Bohr magneton, and B is the applied magnetic field parallel to the wire [12]. One can use the electrostatic environment to control μ, such that one has control over the transition. We will not go into further details about the specific physics of realising MZMs in hybrid devices but instead focus on the numerical work needed to predict suitable designs.

1.4 Motivation to investigate new methods

Hybrid devices are a promising platform for interesting experiments. Without numerical simulations, the only way to design hybrid devices is using intuition and experimental trial and error through multiple rounds of fabrication. With simulations, designs can be adjusted and improved before fabrication starts, making simulations an essential part of the process of designing novel device geometries. The two most widespread methods for numerical simulations of nanoelectronic hybrid devices (and to our knowledge, actually the two only methods being used to simulate real devices) are the Thomas-Fermi (TF) method and the Schrödinger-Poisson (SP) method. These two methods form the basis of this thesis and are extensively discussed in later sections (see section 4.3 for a derivation of the TF method and section 5 for a discussion of the SP method). The TF method is generally a high-speed algorithm, but it does show some unphysical behaviour in certain regimes. In contrast to this, the SP method is very computationally costly but produces high-quality simulations. In this thesis, we will discuss the computational characteristics of both the TF and SP methods, and thus they will be described in relation to each

other. Therefore, when we state, for instance, that the TF method is fast, we mean that it is fast compared to the SP method. When discussing the quality of the simulations, we will take the SP method to be closer to the actual experimental results compared to the TF method. So far, there has been no middle way of doing electrostatic simulations. One either had to pick the high-speed TF method (but somewhat imprecise) or the very slow SP method (but of very high quality). This trade-off is illustrated in Fig. 1.3.

Speed of simulation   Quality of simulation

TF method   ETF method   SP method
- Very fast   - Very slow
- Unphysical in some regimes   - Very high quality
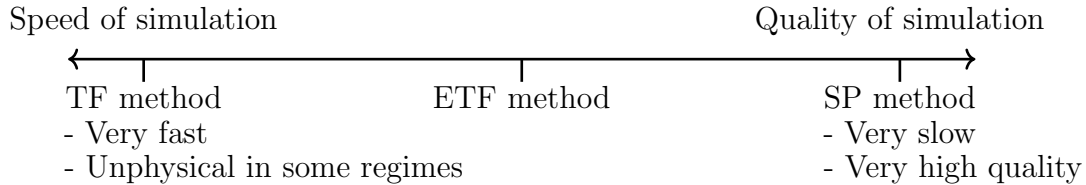
Figure 1.3: Illustration of the speed/quality trade-off in electrostatic simulations. We introduced the ETF method, which lies between the TF and SP methods with respect to speed and quality.

In Fig. 1.3 we see an axis with the speed of the simulation increasing to the left and the quality increasing to the right. The TF method is located in the fast but low-quality end, whereas the SP method is located in the slow but high-quality end. This thesis introduces the extended Thomas-Fermi (ETF) method, which lies somewhere between the TF and SP methods. Thus we bridge the computational gap between the two standard methods with this new method. The ETF method is described in greater detail in section 4.4.

1.5 Outline

This thesis is structured as follows:

- In chapter 2 we review the basics of electrostatics and semiconductor physics, with a particular focus on the parts essential to the simulations discussed in later chapters.

- In chapter 3 we discuss a simplified model of a typical hybrid device and attempt to apply an analytical approach to the evolution of the electric potential through the device.

- In chapter 4 we review the basics of orbital-free density functional theory. We use this to derive the Thomas-Fermi electron density, which will form the foundation for the TF method. Furthermore, we extend the Thomas-Fermi functional with the so-called von Weiszäcker correction and derive an alternative way to solve for the electron density, forming the foundation for the ETF method.

- In chapter 5 we review the numerical methods used in this thesis. On top of this we discuss the Schrödinger problem, that will form the foundation of the SP method in this thesis. We also derive a weak formulation of the ETF functional, which is used to solve for the electron density numerically.

- In chapter 6 we show the results obtained using the three methods mentioned above (SP, TF, and ETF). We compare the methods and apply these to real hybrid devices.

- In chapter 7 we give a conclusion and an outlook, summarising the important results found throughout this thesis and discussing future work.

Chapter 2

Electrostatics and semiconductor physics

In this section, we review the electrostatic problem and the most basic semiconductor physics. This background knowledge will turn out essential for the simulations in this thesis, as it forms the basis for all calculations. We will, however, limit ourselves to the topics that have direct relevance to this thesis, and thus we will not deep-dive into the details of the interesting fields that we encounter.

We will start of by discussing some general topics regarding simulations of hybrid devices.

2.1   Simulating hybrid devices

Hybrid devices are very complex to simulate, as many different branches of physics come into play. We have, for instance, superconductivity, quantum mechanics, and electrostatics. As the computational task of including all aspects that arise in a real device within a simulation is immensely challenging (perhaps often impossible), one has to make approximations. In this thesis, we shall restrict ourselves to describing electrostatics. The electrostatic simulations have a very practical use, as the devices are controlled by voltages applied on its gates. Electrostatic simulations can be used to control that the gate geometry and voltage is working as intended, i.e. the electrostatic environment of the device can be mapped out. The electrostatic environment can, for instance, be important for the location of MZMs and the amount of quantum dots in tunnel barriers, which are extremely relevant for Majorana physics. The goal is thus to understand, describe, and predict electrostatics in hybrid devices. Specifically, we seek to map out the electric potential and the electron density in the devices on a nanometer scale, as these quantities are essential for understanding how to construct a working hybrid device.

In order to make these simulations, we make some basic assumptions regarding the setup of the systems:

- All metals considered are assumed to be perfect conductors. This means that we can fix the electric potential and that the metal will screen all electric fields.

- If not otherwise stated, we will neglect surface and interfacial states. This will be the case for the majority of this thesis, but we will at some points discuss the effect of, for instance, surface charge.

- As is also briefly mentioned above, we will focus purely on electrostatics. This means that we will generally neglect effects such as spin-orbit coupling, magnetic field effects, and superconductivity.

- All simulations assume a low-temperature regime where thermal excitations are much lower than the bandgap, such that thermal excitations can be neglected. This is further described in section 2.4.1.

2.2   The electrostatic problem

To model the electrostatic environment, the electric potential has to be coupled with the charge density. This corresponds to solving the electrostatic problem, which exactly deals with this coupling. Poisson's equation is the cornerstone of electrostatics, as it determines the interplay between charge density and electric potential [4, 16]. Poisson's equation is given by [4, 16]

$$-\boldsymbol{\nabla} \cdot (\epsilon \boldsymbol{\nabla} \phi) = \rho[\phi]. \tag{2.1}$$

Here ϵ is the permittivity of the material, ϕ is the electric potential, and ρ is the charge density. We use the notation $\rho[\phi]$ that specifies that ρ is a functional of the argument in square brackets, i.e. ϕ in this case. We will use this notation to specify functionals in the remainder of this thesis.

For our purpose, we will generally consider the charge density on the form

$$\rho[\phi] = e(p[\phi] - n[\phi]) + \delta(\mathbf{r} - \mathbf{r}_0)\sigma_{\mathrm{s}}[\phi] + \rho_{\mathrm{fixed}}. \tag{2.2}$$

Here e is the electron charge, $p[\phi]$ is the density of holes, $n[\phi]$ is density of electrons, $\delta(\mathbf{r} - \mathbf{r}_0)\sigma_{\mathrm{s}}[\phi]$ is surface charge of the system, and ρ_{fixed} is constant charge. For many cases in this thesis, we will however simplify $\rho[\phi]$ and consider only the electron density.

For all purposes in this thesis, we will use two different boundary conditions for Poisson's equation. We use a Dirichlet boundary condition of $\phi(\mathbf{r}) = V$, where V is a constant electric potential, on all boundaries in contact with gates or wires. For all other boundaries we will use a Neumann boundary condition of $\boldsymbol{\nabla}\phi(\mathbf{r}) \cdot \mathbf{n} = 0$, where $\mathbf{n}$ is the normal vector of the boundary.

Poisson's equation fully describes the electric potential part of the electrostatic problem. However, as the charge density is a functional of the electric potential, the problem has to be solved self-consistently. This can be done by a recursive loop, but in order to close the self-consistent loop, one needs to describe the charge density as a function of electric potential, just as Poisson's equation describes electric potential as a function of electron density. Using such a description of the charge density, a self-consistent system of equations can be formed when combined with Poisson's equation. In this thesis, several different ways of closing the electrostatic loop are investigated. This includes the options underlined below:

- **Analytical models**: Here we describe the evolution of the electric potential through the semiconducting stack using Gauss' law. Accompanied by an ansatz for the charge density, this gives us an analytical expression for the electron density. This is done in section 3.

- **Thomas-Fermi method**: Here we describe the charge density using orbital-free density functional theory and arrive at the Thomas-Fermi electron density. This

is an analytical expression describing the electron density as a function of electric potential. This is derived in section 4.3.

- **Extended Thomas-Fermi method**: Here we again describe the charge density using an orbital-free approach, but here add another term to the Thomas-Fermi energy functional. This added term is the so-called von Weizsäcker term and takes spatial modulation of the electron density into account. This is derived in section 4.4.

- **Schrödinger-Poisson method**: Here we use the Schrödinger equation to solve for the charge density for a given electric potential. This closes the self-consistent loop and allows us to do electrostatic calculations numerically. This is shown in section 5.

The last three methods are used for real device simulations in section 6. Here we seek to provide good electrostatic simulations for hybrid devices and to underline the differences between the methods.

2.3   Semiconductor nanostructures

We define nanostructures as structures with a characteristic length scale between roughly 1 nm and 1 μm [2]. This interval is of course to be understood approximately, as there is no clear transision as to when something is a *nanostructure*. Nanostructures are often also referred to as mesoscopic systems [2]. In this thesis, we deal with semiconducting nanostructures of hybrid devices.

Several different properties characterize semiconductors. As the name suggests, electrical conductivity is one of those properties. Generally speaking, one can say that metals have large conductivities, insulators small, and semiconductors somewhere in between the two [2]. One property of semiconductors (actually a property shared with insulators) that will be especially important in this thesis is the band structure. At zero temperature, the valence band of a semiconductor is completely filled, while the conduction band is completely empty [2]. A bandgap E_g separates these bands, and the Fermi level E_{F} is defined in the middle of this bandgap [2]. This distinguishes semiconductors and insulators from metals, as metals have their conduction bands filled up to the Fermi energy, and their lowest electronic excitations have an arbitrarily small energy cost [2]. One can distinguish between semiconductors and insulators by the size of their bandgap. Typically semiconductors have a bandgap of roughly between 0 eV and 3 eV [2].

2.4   Electronic structure of semiconductors

This section will briefly review the typical band structure of the semiconductors used in this thesis (InAs, InAlAs, and InGaAs). We will take our starting point in the so-called envelope function approximation (EFA), which allows for a comprehensive description of electrons and holes [17]. We will use the simplest approach within EFA

[17], which is the effective-mass approximation. This approximation assumes single, isotropic, parabolic bands. For more advanced approaches, we can refer the interested reader to $\mathbf{k} \cdot \mathbf{p}$ theory, as is for instance described in Ref. [17].

Using the assumption of isotropic, parabolic bands, we will assume a band structure of the form [10]

$$E(\mathbf{k}) = E + \frac{\hbar^2 \mathbf{k}^2}{2m_{\text{eff}}}. \tag{2.3}$$

Here $E(\mathbf{k})$ is the band energy, $\mathbf{k}$ is the wave vector, E is the band energy at $\mathbf{k} = 0$, and m_{eff} is the effective mass. We thus see that the effective mass serves the purpose of modifying the curvature of the parabolic bands. A qualitative sketch of parabolic bands is shown in Fig. 2.1.
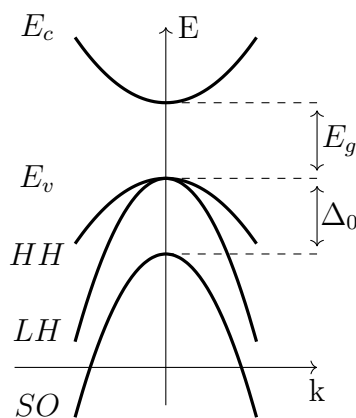
Figure 2.1: Qualitative sketch of the band structure of a direct semiconductor. E_c is the conduction band, E_v is the valence band formed of HH (band for heavy holes), and LH (band for light holes). Figure recreated from [17].

Here E_c is the conduction band, E_v is the valence band, HH is the heavy hole band, and LH is the light hole band. Together the heavy hole band and the light hole band form the valence band. Spin-orbit (SO) interaction gives rise to a splitting of the valence band, shifting this by Δ_0 and forming the band denoted by SO in Fig. 2.1 [17]. For now, we will neglect the SO interactions, but we will briefly mention it at the end of this thesis. The lowest point of the conduction band is called the conduction band minimum, and correspondingly the highest point of the valence band is called the valence band maximum. The energy difference between these two points is the bandgap, E_g.

In Fig. 2.1 we see how the curvatures of the conduction band, the heavy hole band, and the light hole band are all different. The curvature of the bands induces the concept of effective mass. Each electron or hole, moving in its respective band, will behave as if it had a mass different from that of a free electron. As all the semiconductors in this thesis have different band curvature, each semiconductor will have a specific effective mass for electrons, heavy holes, and light holes.

As briefly mentioned above, semiconductors in a low-temperature regime will have a completely filled valence band and a completely empty conduction band. Thus from

the sketch in Fig. 2.1 we see that it would require an energy of the bandgap, E_g, to excite an electron from the valence band to the conduction band. When an electron is in the conduction band, it will be able to conduct electricity, and we will thus consider the electron as free.

Semiconductors can be defined in different groups, according to where their valence band maximum is relative to their conduction band minimum. For instance, the semiconductors GaAs and InAs, as we will use heavily in this thesis, are so-called *direct semiconductors*, as their valence band maximum is in the same place in the Brillouin zone as their conduction band minimum (i.e. they occur for the same wavevector) [2]. In contrast to this, semiconductors like Si and Ge are *indirect semiconductors*, as their valence band maximum and conduction band minimum occur for different wavevectors in the Brillouin zone. The sketched bands in Fig. 2.1 correspond to a direct semiconductor.

2.4.1 Effective band structure for heterostructures

For an isolated semiconductor, we would require an energy of E_g in order to excite an electron from the valence band to the conduction band. However, the semiconductor is in partial contact with aluminium in the simulations described. As aluminium is a metal and thus has many free electrons even at very low temperature, some of those might flow down in the semiconductor's conduction band. We will model this by assigning an offset between the aluminium Fermi level and the conduction band minimum of the semiconductor and call this offset ϕ_w. A sketch of a typical heterostructure in contact with an aluminium wire can be seen in Fig. 2.2.
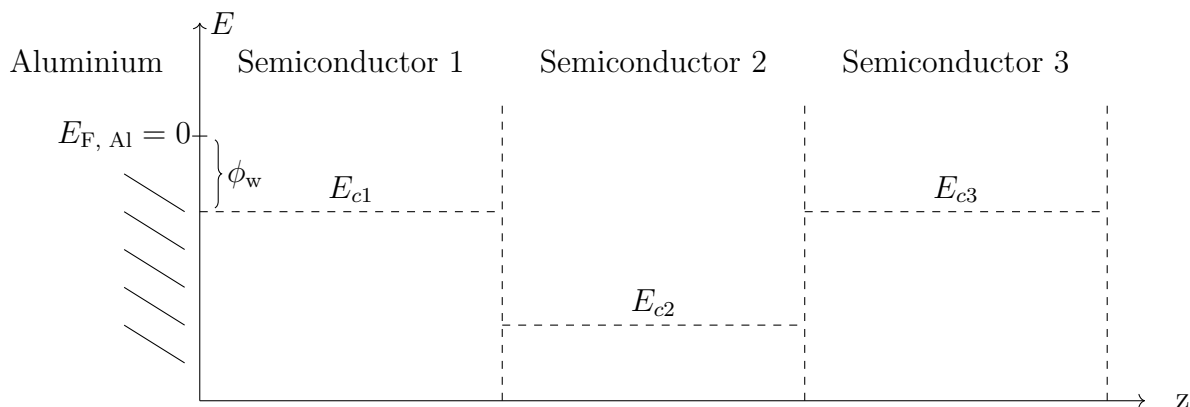
Figure 2.2: Sketch of the convention we use for the band calculations. $V = 0$ corresponds to the Fermi level of aluminium. Here shown without band bending for simplicity.

Here we fixed the Fermi level of aluminium to 0 and thus used this as our reference level. Experimentally one can measure ϕ_w, thus estimating how many electrons flow from the aluminium to the semiconductor at 0 applied voltage. The semiconductor in contact with the aluminium will have a conduction band minimum of E_{c1}, which

will be ϕ_{w} away from $V = 0$, as can be seen in Fig. 2.2. The following semiconductors then have conduction band minimums E_{c2}, E_{c3}, and so on. Every semiconductor will accumulate electrons when the electric potential is larger than the corresponding conduction band minimum. The different semiconducting layers are needed to confine the electrons to a quantum well. The different conduction band minimums form the quantum well that confines the electrons to the 2DEG. In Fig. 2.2 we see how E_{c2} is lower than E_{c1} and E_{c3}, and thus this particular semiconducting stack would confine the electrons primarily to semiconductor 2.

Since we set the zero point to the Fermi level of aluminium, we can easily apply voltages to the system. For instance, applying a voltage of 0.1 V to the aluminium corresponds to filling the semiconductor at the very interface with the aluminium with electrons corresponding to 0.1 eV + ϕ_{w}. This is because we assume that the metal-semiconductor interface is an Ohmic contact, i.e. there is no barrier at the interface. As we start filling the semiconductor bands, we naturally start to see band bending, which is shown in the results below.

Note that we completely neglect the semiconductor's valence band. This is only a good approximation at temperatures where the thermal excitations of the valence electrons are much smaller than the bandgap, i.e. $k_{\mathrm{B}}T \ll E_g$, where k_{B} is the Boltzmann constant, T is the temperature, and E_g is the bandgap between the valence band maximum and the conduction band minimum [10]. However, the experiments are conducted at temperatures on the sub-kelvin scale, and thus we are in a regime where these thermal excitations can be neglected.

Chapter 3

Simple device modelling

This thesis considers several different hybrid device designs. This naturally happens as experimental physicists vary their geometries in search of better results, and thus theoretical models and simulations of different geometries are essential. For all the geometries considered in this thesis, the basic physical configurations vary only slightly. For our modelling purpose, we explore two different regions. The first region in focus is the metal-oxide-semiconductor-region (MOS-region), and the second is the wire-semiconductor-region (wire-region). These regions are illustrated in Fig. 3.1.
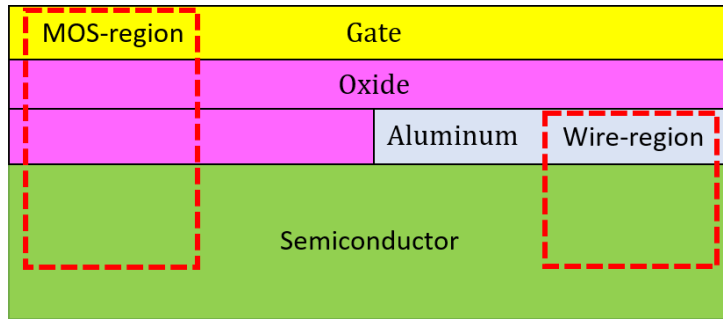
Figure 3.1: Typical cross section of the devices investigated in this thesis. MOS-region (left) and wire-region (right) illustrated with red dotted lines. The semiconducting stack (green) is a heterostructure that forms a quantum well.

In Fig. 3.1 we see a cross-section of a typical device. We have a metal gate (yellow) that is separated from an aluminium wire (grey) by a layer of oxide (purple). Below this, we have a semiconducting stack (green). The MOS-region is defined on the side without the wire, and the wire-region is defined on the side including the wire, as can be seen on the two red boxes in Fig. 3.1.

3.1   MOS-region

The metal part of the MOS-region consists of a metallic gate (Ti/Au) with an applied voltage, V_G. The following part is the oxide (HfO_2). This is followed by the semiconductor part, which physically consists of a barrier and a well. This is summarised in Fig. 3.2.
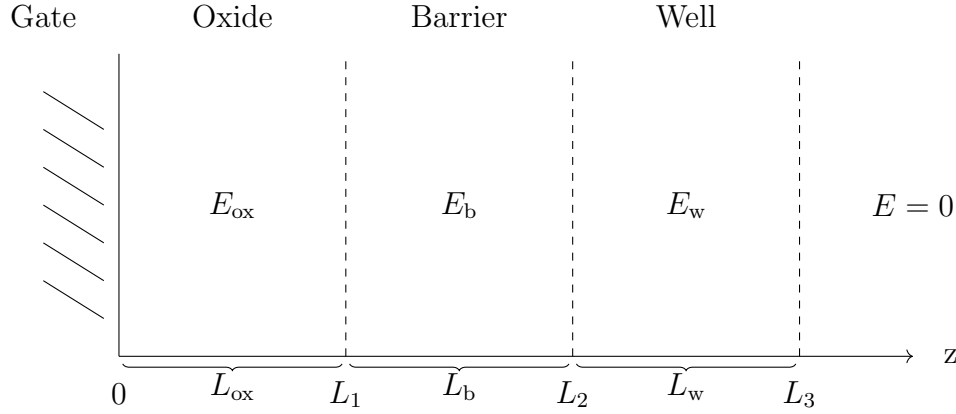
Figure 3.2: Sketch of the simple 1D model.

In Fig. 3.2 we see how the oxide, the barrier, and the well, are given lengths L_{ox}, L_{b}, and L_{w}, respectively, along with electric fields E_{ox}, E_{b}, and E_{w}. For this simple model, we assign a Neumann-boundary condition after the well by fixing the electric field to zero, $E = 0$. With the stack shown in Fig. 3.2, we will assume a uniform system in x- and y-directions, and thus we are effectively solving a 1D problem. We apply the following to this model:

- A constant volume charge in the oxide, $\rho_{\text{ox}} = \text{const}$. In real devices, we expect this charge to be caused by positive ions.

- An interface surface charge between the oxide and the barrier, $\sigma_s = e\, g_{\text{DIT}}\big(-eV_{\text{s}} + \phi_{\text{s}}\big)$. Here e is the electron charge, g_{DIT} is the 2D density of states at the interface, V_{s} is the voltage at the interface, and ϕ_{s} is the Fermi-energy mismatch. Physically we expect this charge to be caused by positive ions.

- A charge distribution in the well, $\rho_{\text{w}} = \sigma_{\text{w}} U(z)$. Here σ_{w} is the charge, and $U(z)$ is the normalized distribution of the charge. We have explored several different distributions of charge, including all the charge localized to one plane (placed in the top, bottom, or centre of the well), a uniform distribution through the well, as well as a $\sin^2$ distribution in the well. All these distributions are summarised in Fig. 3.3.

$$U(z) = \frac{2}{L_\mathrm{w}} \sin \left(\pi z / L_\mathrm{w} \right)^2$$
$$\alpha = 1/3 + 5/(8\pi^2) \approx 0.4$$

$$U(z) = 1/L_\mathrm{w}$$
$$\alpha = 1/3$$

$$U(z) = \delta(z)$$
$$\alpha = 0$$

$$U(z) = \delta(z - L_\mathrm{w}/2)$$
$$\alpha = 1/2$$

$$U(z) = \delta(z - L_\mathrm{w})$$
$$\alpha = 1$$
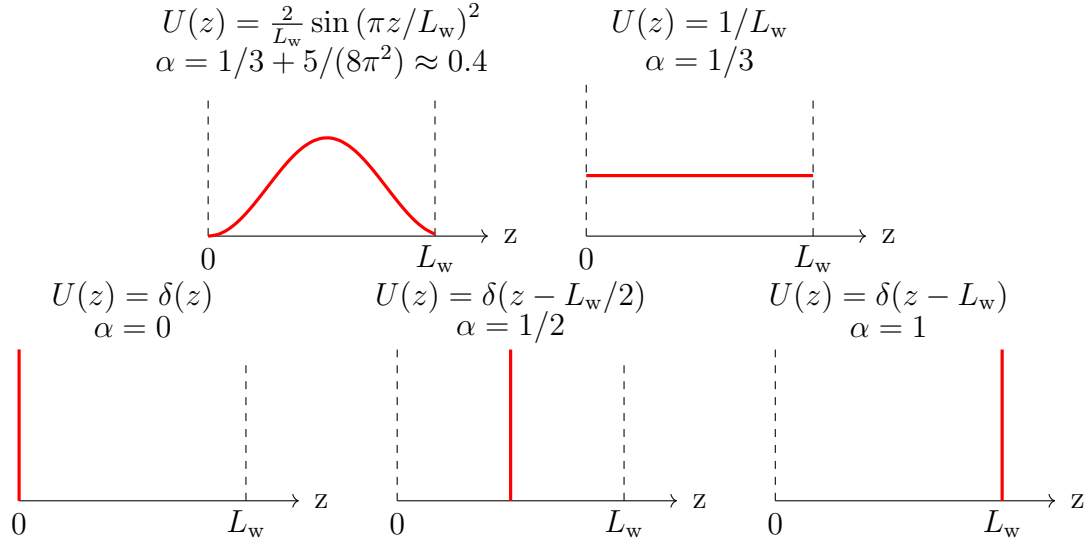
Figure 3.3: Sketches of the five different charge distributions in the well.

In the above, we assume σ_w to be of the form

$$\sigma_\mathrm{w} = -e \, g_\mathrm{w}(-\varepsilon + \phi_\mathrm{w}). \tag{3.1}$$

Here g_w is the 2D density of states, ε is the electrostatic energy for an electron in the well, and ϕ_w is the Fermi energy mismatch. Experimentally we assume this charge to be caused by electrons, but this is not assumed in the following.

Using Gauss' law [16], we can describe how the electric field evolves through the system in Fig. 3.2. Gauss' law states that [16]

$$\int_{\partial \Omega} \epsilon \mathbf{E} \cdot \mathbf{dS} = \int_{\Omega} \rho \, d\mathbf{r}. \tag{3.2}$$

Here Ω is the domain of integration, and $\partial \Omega$ is the edge of this. $\mathbf{dS}$ is the normal of an infinitesimal surface element of the domain, ϵ is the permittivity, $\mathbf{E}$ is the electric field, and ρ is the charge density. Naturally, Gauss's law in Eq. (3.2) can be immensely hard to analytically use in real systems. The clever thing about the simplified MOS model in Fig. 3.2 is that Gauss's law can easily be applied here (since the MOS model is effectively 1D). Thus we hope to gain insight into the electrostatics of the real systems by analysing this simplified model. In the case of a simplified 1D system, the electric potential is given by [16]

$$V = -\int E \, dz, \tag{3.3}$$

where z is along the direction shown in Fig. 3.2. As we can describe the electric field by Gauss's law in Eq. (3.2), we can now describe the electric potential as well. Using that the electrostatic energy, ε is given by

$$\varepsilon = -e \int_0^{L_\mathrm{w}} U(z)V(z)dz, \tag{3.4}$$

where $V(z)$ is the electric potential in the well as a function of the spatial variable z, we can analytically solve for the electron density in the well. Inserting the expression for ε from Eq. (3.4) in the expression for σ_w in Eq. (3.1), and then solving for the electron density, $n_w = \sigma_w/(-e)$, gives us (calculation is shown in appendix A)

$$n_w = \frac{1}{e}\left(\frac{1}{e^2 g_w} + \alpha\frac{L_w}{\epsilon_w} + \frac{L_{ox}}{\epsilon_{ox}} + \frac{L_b}{\epsilon_b}\right)^{-1}\left(V_G + \frac{L_w}{\epsilon_w}(\sigma_s + \rho_{ox}L_{ox}) - \frac{\phi_w}{e}\right). \tag{3.5}$$

Here ϵ_{ox}, ϵ_b, and ϵ_w are the permittivities of the oxide, the barrier, and the well, respectively. We also introduced the parameter α, which turns out to be the only parameter that describes the electron distribution in the well. With an electron distribution localized as a Dirac delta function in the front end of the well, we have $\alpha = 0$. With an electron distribution localized in the back of the well, we have $\alpha = 1$, and with an electron distribution localized in the middle of the well, we have $\alpha = 1/2$. For the uniform case, we get $\alpha = 1/3$, and for the $\sin^2$ distribution, we have $\alpha = 1/3 + 5/(8\pi^2) \approx 0.4$. This is summarised in Fig. 3.3.

This means that varying the electron distribution inside the well applies a weight α to the inverse capacity of the well, L_w/ϵ_w. Physically this makes sense, as the well effectively is made infinitesimally small when all the charge is localized in the front end, and we thus have a weight of $\alpha = 0$ here. When the charge is localized in the back end, we have a maximum weight of $\alpha = 1$. Thus all distributions inside the well will only matter with a weight between $\alpha = 0$ and $\alpha = 1$.

From Eq. (3.5) we can conclude that the electron density n_w is linear in gate voltage. Another conclusion we can draw from Eq. (3.5), is that the dependence of gate voltage is not affected by the constant charge in the oxide, ρ_{ox}, as this enters as a constant offset. Thus ρ_{ox} physically lifts (or lowers) the electron density in the well, but it does not affect the sensitivity to gate voltage. We will now explore whether these conclusions still hold if $U(z)$ is not solely defined in the well.

3.1.1 Charge not localized in the well

We will now allow the electron distribution $U(z)$ to leave the well and enter both the barrier and the oxide. Using Gauss' law [16], we get the following expression for the electric field in the well

$$E_w = E_0 + \frac{\sigma_{ox}^{int}}{\epsilon_{ox}} + \frac{\sigma_s}{\epsilon_{ox}} + \frac{\sigma_b^{int}}{\epsilon_b} + \frac{\sigma_w^{int}}{\epsilon_w}, \tag{3.6}$$

where E_0 is the electric field at $z = 0$, and we have defined

$$\sigma_{ox}^{int} = \int_0^{L_1} \rho_w dz + \int_0^{L_1} \rho_{ox} dz, \quad \sigma_b^{int} = \int_{L_1}^{L_2} \rho_w dz, \quad \text{and} \quad \sigma_w^{int} = \int_{L_2}^{L_3} \rho_w dz. \tag{3.7}$$

Using this, we can calculate the electric potential in the well

$$V_w = V_G - \int_0^{L_3} E_w dz, \tag{3.8}$$

where V_G is again the voltage applied at the gate. This gives us the electrostatic energy per electron

$$\varepsilon = -e \int_0^{L_3} U(z) V_w \mathrm{d}z. \tag{3.9}$$

Recalling that we assumed $\sigma_w = -e \, g_w(-\varepsilon + \phi_w)$, we can tell from Eq. (3.9), that ρ_{ox} (when assumed constant) does not affect the slope of the electron density, with respect to gate voltage. This is clear as ρ_{ox} only enters through Eq. (3.7), but here only as a constant. Since it contributes only by being added here, and not by being multiplied, we see that even for a general $U(z)$, the slope of the electron density is not affected by ρ_{ox}.

3.1.2 Calibration from Hall bar measurements

Considering now the predicted electron density n_w in Eq. (3.5), it predicts a density sensitivity to voltage of

$$\frac{dn_w}{dV_G} = \frac{1}{e} \left(\frac{1}{e^2 g_w} + \alpha \frac{L_w}{\epsilon_w} + \frac{L_{ox}}{\epsilon_{ox}} + \frac{L_b}{\epsilon_b} \right)^{-1}. \tag{3.10}$$

Using a Hall bar measurement (see for instance Ref. [18] for further details on Hall bar measurements), one can obtain detailed information about the electron density in the semiconducting stack as a function of gate voltage. Data from a Hall bar measurement can be seen in Fig. 3.4.
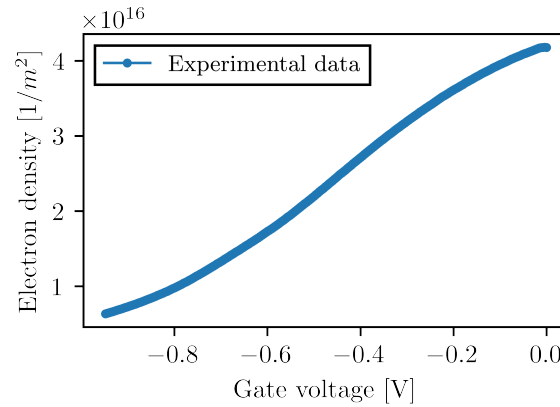
Figure 3.4: Hall bar data of the electron density as a function of gate voltage. For more details about fabrication see Ref. [19], and for a paper using the very same stack, see Ref. [20].

The data presented in Fig. 3.4 is from a semiconducting stack consisting of an InAs well ($L_w = 7$ nm, $\epsilon_w = 15.15 \, \epsilon_0$), a HfO_2 barrier ($L_{ox} = 15$ nm, $\epsilon_{ox} = 25 \, \epsilon_0$), and an $In_{75}Ga_{25}As$ barrier ($L_b = 10$ nm, $\epsilon_b = 14.76 \, \epsilon_0$), where ϵ_0 is the vacuum permittivity.

Using this data, we want to test the simplified model, in particular the electron density's sensitivity to gate voltage. In Fig. 3.4 we see that the measured electron

density is not exactly linear in gate voltage, but a straight line seems to be a decent fit, as the simple model predicts. In Fig. 3.5 we did a linear fit on the experimental data (orange graph), and plotted a straight line with the predicted slope from Eq. (3.10), taking the same starting point as the linear fit (green graph). One should thus only consider the slope of the green graph. For this calculation we used $g_{\text{w}} = m_{\text{eff}}/\pi\hbar^2$, where the effective mass is taken to be the one of InAs, i.e. $m_{\text{eff}} = 0.026\ m_{\text{e}}$, and we have used $\alpha = 1$. This particular value of g_{w} is justified by the Thomas-Fermi approximation, which is further discussed in section 4.3.
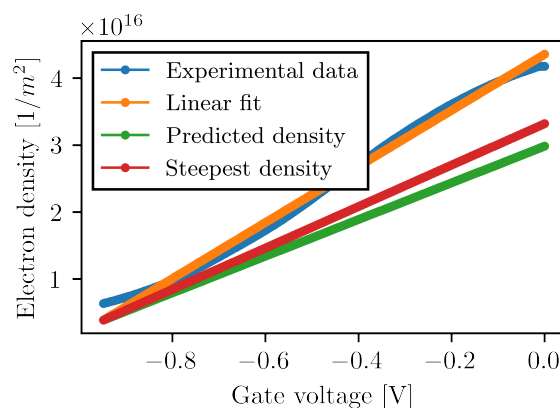
Figure 3.5: Hall bar data of the electron density as a function of gate voltage (blue graph). Linear fit of the experimental data (orange graph), along with the predicted electron density from Eq. (3.10) (green graph), as well as the predicted electron density with $\alpha = 0$ (red graph).

As can be clearly seen In Fig. 3.5, the slope of the predicted density is not similar to the slope of the linear fit. However, we used $\alpha = 1$ in this calculation, and one could argue for a physical system with $\alpha = 0$. We have plotted the slope from Eq. (3.10) with $\alpha = 0$ on the red graph in Fig. 3.5.

This value for α should be corresponding to the largest slope one can get from adjusting the electron distribution in the well. However, we see that even this steepest density is still not as steep as the linear fit. One could argue that perhaps we are not just filling one band and that g_{w} should be multiplied by some integer. It turns out that even making g_{w} five times larger still does not give a slope as steep as the linear fit. Thus we do not seem to be able to explain the sensitivity of the electron density with respect to gate voltage using the simplified model. Therefore we have to rethink the method and go to more advanced descriptions of the electron density.

3.2 Wire-region

The second region we will consider is the wire-region. As all the hybrid devices considered in this thesis contain some sort of superconducting wire, we have to consider how the superconductivity affects the electric potential and electron density. For all devices

considered in this thesis, the superconducting parts will lay on top of the semiconducting layer, and thus we have a superconductor-semiconductor interface. One should take hybridization into account here; see for instance Ref. [4]. For now, however, we will focus on electrostatics. This means that we will model the superconducting wire as a normal metal, thereby neglecting superconductivity. Thus, many of the MOS-region considerations apply here to the approximate modelling of the wire as a simple metal. However, with the metal-semiconductor interface, other problems have to be addressed. We will assume the interface to be an Ohmic contact with a band offset between the Fermi level of the metal and the conduction band minimum of ϕ_{w}.

Chapter 4

Orbital-free density functional theory

The experimental systems explored in this thesis are quantum mechanical systems consisting of N electrons affected by some external potential. Such systems can be solved for the corresponding many-body wave function of dimensionality $3N$ [21]. Even without considering spin, optimization of the wave function quickly becomes computationally expensive [21, 22]. Therefore a computationally cheaper approach is of great interest. This is further described in the sections below.

4.1   Introduction to density functional theory

Intuitively, one could think that since electrons are indistinguishable, it might be enough to consider the total electron density, i.e. the probability to find *any* electron at a given point, rather than calculating the entire system for *all* the electrons [21]. This is indeed possible, and it is the line of thought in orbital-free (OF) density functional theory [21]. As is further described below, OF density functional theory takes the electron density as the basic variable and removes the need to solve the Schrödinger eigenvalue problem. This changes the problem's dimensionality from $3N$ to 3 [21], which greatly speeds up the numerical calculations.

In OF density functional theory, we generally consider a system of N electrons in an external potential V_ext, where the total energy E can be expressed as a functional of the electron density $\rho(\mathbf{r})$ [21]. The energy then takes the following form [21]

$$E[\rho] = F[\rho] + \int V_\text{ext}\rho(\mathbf{r})d\mathbf{r}. \tag{4.1}$$

Here the integration is over the entire volume of the system, and $F(\rho)$ is some universal function that contains all information about how the electrons interact and behave [21]. The exact form of $F(\rho)$ is unknown, so one has to approximate this to optimize the energy functional. Traditionally, and as we shall do later on in this thesis, $F(\rho)$ is split into a kinetic and a potential part.

4.2   General orbital-free functional theory

Above we have made some simplifications in the functional. When applied to real devices, we will have a complicated electric field dependence, and thus the functional will no longer be a functional of ρ only, but also of the electric potential ϕ. This generally makes the problem harder to solve, as it is no longer just a density functional.

The full, coupled functional for *any* system is generally given by [23]

$$E[\rho, \phi] = E_\text{k}[\rho] + \int V_\text{ext}\rho(\mathbf{r})d\mathbf{r} + \int \rho(\mathbf{r})\phi(\mathbf{r})d\mathbf{r} + \int E_\text{EM}[\phi]d\mathbf{r}. \tag{4.2}$$

Here $E_k[\rho]$ is the ρ-dependant kinetic energy functional, V_{ext} is an external potential, and $E_{\text{EM}}[\phi]$ is the electric field energy given by $\epsilon|\boldsymbol{\nabla}\phi(\mathbf{r})|^2/2$.

As the functional in equation (4.2) is very complicated, we need to simplify it before we can minimise it. We will do this by considering free charges without complicated boundary conditions, and thus prove that in this regime, we can derive the simpler functional in equation (4.1). If we minimise the functional in equation (4.2) with respect to $\phi(\mathbf{r})$ we see that we get [23]

$$\frac{\delta E}{\delta \phi} = \rho(\mathbf{r}) + \boldsymbol{\nabla} \cdot (\epsilon \boldsymbol{\nabla} \phi(\mathbf{r})) = 0, \quad \rightarrow \quad -\boldsymbol{\nabla} \cdot (\epsilon \boldsymbol{\nabla} \phi(\mathbf{r})) = \rho(\mathbf{r}), \qquad (4.3)$$

which is exactly Poisson's equation. Assuming a constant permittivity leads to the simplified version [23]

$$-\boldsymbol{\nabla}^2 \phi(\mathbf{r}) = \rho(\mathbf{r})/\epsilon. \qquad (4.4)$$

Using that the fundamental solution to the problem

$$-\boldsymbol{\nabla}^2 G(\mathbf{r}) = \delta(\mathbf{r}), \qquad (4.5)$$

where $\delta(\mathbf{r})$ is the Dirac's delta function, is given by [23]

$$G(\mathbf{r}) = \frac{1}{4\pi|\mathbf{r}|}, \qquad (4.6)$$

we can get an expression for ϕ in Poisson's equation. We see that

$$\phi(\mathbf{r}) = \int G(\mathbf{r} - \mathbf{r}')\rho(\mathbf{r}')d\mathbf{r}' = \int \frac{\rho(\mathbf{r}')}{4\pi\epsilon} \frac{1}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}'. \qquad (4.7)$$

However, equation (4.6) is only an exact solution in free space. For real devices, there are complicated boundary conditions and changing permittivities, and thus this solution is not correct. However, keeping this solution means that we can approximate $\int \rho(\mathbf{r})\phi(\mathbf{r})d\mathbf{r}$ from equation (4.2) as

$$\int \rho(\mathbf{r})\phi(\mathbf{r})d\mathbf{r} = \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{4\pi\epsilon} \frac{1}{|\mathbf{r} - \mathbf{r}'|}. \qquad (4.8)$$

With this simplification, we can write the total energy functional as a functional of ρ only

$$E[\rho] = E_k[\rho] + \int V_{\text{ext}}\rho(\mathbf{r})d\mathbf{r} + \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{4\pi\epsilon} \frac{1}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}', \qquad (4.9)$$

as we have seen that the functional minimisation with respect to ϕ directly gives us Poisson's equation. This means that we can now minimise the functional with respect to ρ, and couple this with Poisson's equation to capture both the physical behaviour from ϕ and ρ. Thus we have *decoupled* the problem and can now write the functional with respect to ρ only.

This is exactly the form that we implicitly assume in the calculations below, and thus the functional we use to derive the TF and ETF methods. This is exact in free

space, as it is under the assumption of ϵ being constant and under the assumption of no boundary conditions. However, this is not correct in real devices. There might be inhomogeneity of the permittivity, and the metals induce non-trivial boundary conditions to the system. Including these features in a density functional theory would make the fundamental solution to Poisson's equation (i.e. for the case discussed above, the solution in equation (4.6)) extremely complicated, which would thus make the interaction term extremely complicated. We will later return to the ρ and ϕ coupled orbital-free functional. Nevertheless, in the next sections, we will use the density functional formulation to discuss some previous results and make a clear connection to the literature [23].

4.3   Thomas-Fermi functional

The SP method described later on numerically solves the Schrödinger equation shown in Eq. (5.7). This approach can be excellent for very small systems. However, as the wave function Ψ is indeed a many-body wave function (for N particles we have $\Psi(\mathbf{r}_1, ..., \mathbf{r}_N)$), the Schrödinger equation becomes immensely complicated to solve. This is caused by the need to diagonalize the Hamiltonian, which is numerically a very expensive operation. However, for many purposes, one can get around this problem by, instead of calculating the many-body wave function directly, focusing on the particle density $n(\mathbf{r})$. For our purpose, the wave function itself is not of practical interest, but the electron density, which is actually a physical observable, is of great interest. The connection between the many-body wave function ($\Psi(\mathbf{r}_1, ..., \mathbf{r}_N)$) and the density ($n(\mathbf{r})$) is the following [24],

$$n(\mathbf{r}) = N \int \Psi^*(\mathbf{r}, \mathbf{r}_2, ..., \mathbf{r}_N)\Psi(\mathbf{r}, \mathbf{r}_2, ..., \mathbf{r}_N)d\mathbf{r}_2...d\mathbf{r}_N. \qquad (4.10)$$

Here N is the total particle number, and $n(\mathbf{r})$ is the probability amplitude for finding a particle at position $\mathbf{r}$ [24]. The Thomas-Fermi (TF) method does exactly this: It bypasses the many-body wave function and treats the density as a basic variable which is solved for.

The TF method is a semi-classical method. For this semi-classical approach to be true, we will assume the de Broglie wavelength, $\lambda(x) = 1/k(x)$, where $k(x) = 1/\hbar p(x)$, to be slowly varying in space. Here we take x to be any spatial coordinate. We also introduced the wavenumber $k(x)$ and the momentum $p(x)$, all with spatial dependence [24]. This gives us the slowly varying condition [24],

$$\left| \frac{d\lambda(x)}{dx} \right| \ll 1. \qquad (4.11)$$

In contrast to quantum mechanics, one should note here that we treat position and momentum classically, i.e. as scalars and not as operators. From quantum mechanics, we will adopt two concepts. Firstly we will use Fermi statistics, which dictate that all states up to a maximum energy $E_{\max}$ with corresponding Fermi momentum p_F are filled. Then, we will use the uncertainty principle, which states that every cell of area

h^2 in phase space (taking here the 2D version, but one can easily generalize to 3D) may contain up to two electrons with opposite spin [24]. h is here Planck's constant. Counting all the electrons in a system of area A, we get that the total number of electrons is:

$$\pi p_F^2 \cdot 2 \cdot \frac{A}{h^2} = N. \tag{4.12}$$

Here πp_F^2 is the contribution from momentum space, the factor of 2 is from the spin, and A/h^2 is the number of cells in the area A. The uniform electron density is thus $n = N/A$. Isolating for the Fermi momentum we get

$$p_F = \sqrt{\frac{h^2}{2\pi}n}. \tag{4.13}$$

Considering now the classical energy for the fastest electron [24], we get that

$$E_{\max} = \frac{p_F^2}{2m} + V = \frac{h^2 n}{4\pi m} + V, \tag{4.14}$$

where we inserted Eq. (4.13) in the last step, and introduced the potential V and the electron mass m. One should note here that, generally speaking, the electron mass and the potential can have a spatial dependence. This does not play a role here, and the spatial dependence is thus suppressed for simplicity. We allow both n and p_F to be spatially dependent, but only such that the semi-classical condition in (4.11) holds. Considering now the kinetic energy density, $t(\mathbf{r})$, such that the total kinetic energy T can be recovered by integrating the density over all space, $T = \int t(\mathbf{r})d\mathbf{r}$, we get [24]

$$t = \frac{T}{A} = \frac{1}{A} \int \frac{p^2}{2m} dN = \frac{1}{A} \int_0^{p_F} \frac{p^2}{2m} \frac{4\pi A p}{h^2}. \tag{4.15}$$

In the last step, we changed integration variable by expanding from dN to $\frac{dN}{dp}dp$, and using the expression for N in Eq. (4.12). Simply calculating this integral gives us [24]

$$t = \frac{2\pi}{mh^2} \int_0^{p_F} p^3 = \frac{\pi}{2mh^2} p_F^4. \tag{4.16}$$

Now using the expression for p_F from Eq. (4.14), we get

$$t = \frac{\pi}{2mh^2} \left(\sqrt{\frac{h^2}{2\pi}n}\right)^4 = \frac{\pi}{2mh^2} \frac{h^4}{2^2\pi^2} n^2 = \frac{h^2}{8\pi m} n^2. \tag{4.17}$$

For simplicity, we will make the definition $C_K \equiv h^2/8\pi m$. Focusing now on the total potential energy of the system, caused by the externally applied potential, V_{ext}, as well as by the Coulomb interaction of the electrons, we get the following [16, 24]

$$U = e \int n(\mathbf{r}) V_{\text{ext}}(\mathbf{r}) d\mathbf{r} + \frac{1}{2} e^2 \int \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}'. \tag{4.18}$$

From this we can express the total energy of the system as $E_{\text{tot}} = T + U$. We get that

$$E_{\text{tot}} = \int \left(C_K n(\mathbf{r})^2 + e n(\mathbf{r}) V_{\text{ext}}(\mathbf{r}) + \int \frac{1}{2} e^2 \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \right) d\mathbf{r}. \tag{4.19}$$

The approach is now to minimise the total energy of the system, under the condition of normalization, $\int n(\mathbf{r})d\mathbf{r} = N$. We will do so by introducing the Lagrange multiplier μ, and vary with respect to $n(\mathbf{r})$ [24]. We thus get that we are to solve the optimization,

$$\frac{\delta(E_{\text{tot}} - \mu N)}{\delta n(\mathbf{r})} = 0. \tag{4.20}$$

This gives us the Lagrangian function

$$\mathcal{L} = \int C_{\text{K}} n(\mathbf{r})^2 d\mathbf{r} + e \int n(\mathbf{r}) V_{\text{ext}}(\mathbf{r}) d\mathbf{r} + \frac{1}{2} e^2 \int \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' - \mu N. \tag{4.21}$$

Using that $\int n(\mathbf{r})d\mathbf{r} = N$, we can rewrite Eq. (4.21) as

$$\mathcal{L} = \int \left(C_{\text{K}} n(\mathbf{r})^2 + e n(\mathbf{r}) V_{\text{ext}}(\mathbf{r}) + \frac{1}{2} e^2 \int \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' - \mu n(\mathbf{r}) \right) d\mathbf{r}. \tag{4.22}$$

Using the kinetic and potential energy densities, t and v, defined such that $E_{\text{tot}} = \int t + v \, d\mathbf{r}$, we see that Eq. (4.22) is simply

$$\mathcal{L} = \int \left(t + v - \mu n(\mathbf{r}) \right) d\mathbf{r}. \tag{4.23}$$

Since the Lagrangian function does *not* depend on the derivative of $n(\mathbf{r})$, the variation with respect to $n(\mathbf{r})$ is simply the derivative of the integrand [25]. We thus get that [25]

$$\frac{\delta \mathcal{L}}{\delta n(\mathbf{r})} = \frac{d(t + v - \mu n(\mathbf{r}))}{dn(\mathbf{r})} = 2 C_K n(\mathbf{r}) + e V_{\text{ext}}(\mathbf{r}) + e^2 \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' - \mu = 0. \tag{4.24}$$

Here we obtain the 2D Thomas-Fermi equation, which determines the equilibrium distribution of the electron density [24]. Indeed the variation of $n(\mathbf{r})$ actually leads to the Euler-Lagrange equation, $\delta E_{\text{tot}}/\delta n - \mu = 0$ [26]. Explicitly we get

$$2 C_K n(\mathbf{r}) + e V_{\text{ext}}(\mathbf{r}) + e^2 \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' = \mu, \quad C_K = h^2/8\pi m. \tag{4.25}$$

Here we note, that if we write Eq. (4.14) with the definition of C_K, we get

$$E_{\text{max}} = \frac{h^2 n(\mathbf{r})}{4\pi m} + V(\mathbf{r}) = 2 C_K n(\mathbf{r}) + V(\mathbf{r}), \tag{4.26}$$

which compared to Eq. (4.25), reveals that $E_{\text{max}} = \mu$, since the two last terms in Eq. (4.25) are potential energy terms. From Eq. (4.24) we also see that $\mu = \partial E_{\text{tot}}/\partial N$ [24]. This means that μ is to be interpreted as the chemical potential of the system. Introducing now the electrostatic potential caused by $n(\mathbf{r})$ as [24]

$$u(\mathbf{r}) = e \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}', \tag{4.27}$$

and for simplicity, the total potential energy $V(\mathbf{r})$ as

$$V(\mathbf{r}) = eV_{\text{ext}}(\mathbf{r}) + eu(\mathbf{r}), \tag{4.28}$$

we can rewrite Eq. (4.25). We get that the density is

$$n(\mathbf{r}) = \frac{1}{2C_K}\Big(\mu - V(\mathbf{r})\Big) = \frac{4\pi m}{h^2}\Big(\mu - V(\mathbf{r})\Big), \tag{4.29}$$

or in terms of $\hbar$:

$$n(\mathbf{r}) = \frac{m}{\pi\hbar^2}\Big(\mu - V(\mathbf{r})\Big). \tag{4.30}$$

This is the electron density from the 2D Thomas-Fermi equation. One can do a similar derivation for the 3D case, which will yield the result [24]

$$n_{\text{3D}}(\mathbf{r}) = \frac{8\pi}{3}\left(\frac{2m}{h^2}\right)^{3/2}\Big(\mu - V(\mathbf{r})\Big)^{3/2} = \frac{1}{3\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}\Big(\mu - V(\mathbf{r})\Big)^{3/2}. \tag{4.31}$$

With Eq. (4.31) we are now able to calculate the electron density corresponding to a given electric potential within the TF model. Thus we see that Eq. (4.31) serves the purpose of coupling the electric potential to the electron density. Therefore we can sketch an algorithm for electron distribution calculations, using Eq. (4.31). We will name this approach the TF method, as it uses the Thomas-Fermi functional to calculate the charge density, which can be used with Poisson's equation in order to close the electrostatic problem self-consistently. It is outlined in Fig. 4.1.
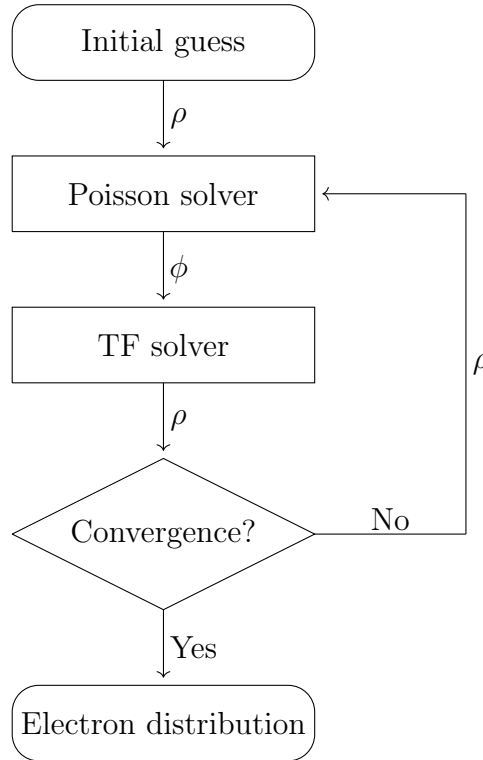
Figure 4.1: Sketch of the Thomas-Fermi-Poisson method (TF method). Starting from an initial guess, the electron distribution of the system is numerically calculated.

Here the TF solver box in the flow chart in Fig. 4.1 refers to the use of Eq. (4.31). The flow chart in Fig. 4.1 shows a segregated approach, i.e. Poisson's equation and the TF electron density are solved one at a time and combined iteratively until convergence. This is opposed to the fully coupled approach where the coupled problem is solved, i.e. $-\boldsymbol{\nabla} \cdot \epsilon \boldsymbol{\nabla} \phi = n[\phi]$. One could easily here use the fully coupled approach instead of the segregated.

In the derivation above, we introduced μ as a Lagrangian multiplier to fix the number of particles. In order to use these results, we need a physical understanding of μ, such that we can apply the above analysis to physical systems.

As μ fixes the number of particles and as can be seen in Eq. (4.31) implements the requirement that $V < \mu$ in order for us to have a non-zero density, we can interpret μ as minus the conduction band minimum (E_c) of the given material. This is because we expect to fill with electrons for potential greater than E_c, and we expect depletion for potentials less than E_c. This naturally follows from the discussion of the semiconductor band structure in section 2.4. For the band calculations, we will use the notation and convention as is sketched in Fig. 2.2 [27], and use $\mu = -E_c$.

4.4 Von Weizsäcker functional correction

From the TF electron densities in 2D and 3D (Eq. (4.30) and (4.31) respectively), we see that the density completely vanishes at $V(\mathbf{r}) = \mu$, and that the density is taken to be zero where $V(\mathbf{r}) > \mu$. This sharp transition in density corresponds to an unphysical behaviour [26], caused by the local nature of the TF approximation. This unphysical behaviour can be removed (or at least improved), by enhancing the TF kinetic energy functional [26]. Furthermore it is shown in Ref. [26], that the TF kinetic energy functional *always* produces an energy lower than the true value. Thus the TF kinetic energy functional should be augmented by some term that raises the kinetic energy and has a non-local behaviour, such that we smooth out the jump in density at $V(\mathbf{r}) = \mu$. One way to fix this is to add the von Weizsäcker (vW) kinetic energy functional [22, 26, 28, 29] to the TF functional. The vW correction takes the form [26]

$$E_{\text{vW}} = \lambda_{\text{vW}}(N) \frac{\hbar^2}{8m} \int \frac{|\boldsymbol{\nabla} n(\mathbf{r})|^2}{n(\mathbf{r})} d\mathbf{r}. \tag{4.32}$$

Here we introduced the gradient of the spatial density, $\boldsymbol{\nabla} n(\mathbf{r})$, and the vW parameter $\lambda_{\text{vW}}(N)$, which generally can depend on the particle number N [26]. However, we will consider it a constant for our purpose and thus suppress the N dependence. Below we will briefly discuss the desired properties of the vW correction in Eq. (4.32). The particular form is further justified in Ref. [28]. We note that the vW correction has the following properties [26]:

- It depends on the gradient of the spatial density, and thus the correction vanishes for a perfectly uniform system, as we would expect (since there the TF approximation is exact).

- It is positive definite and thus adds energy to the TF approximation, which has energy lower than the true value.

- It is gradient dependent and therefore induces a short-range non-locality compared to the completely local TF method.

Using this correction in the TF energy functional gives us the total energy functional (TF part from Eq. (4.19))

$$E_{\text{tot}} = \int \left(\frac{\hbar^2 \pi}{2m} n(\mathbf{r})^2 + \lambda_{\text{vW}} \frac{\hbar^2}{8m} \frac{|\boldsymbol{\nabla} n(\mathbf{r})|^2}{n(\mathbf{r})} + en(\mathbf{r}) V_{\text{ext}}(\mathbf{r}) + \frac{1}{2} e^2 \int \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \right) d\mathbf{r}. \quad (4.33)$$

Thus we have extended the TF energy functional with the vW correction and obtained the *extended Thomas-Fermi* (ETF) energy functional. The naming here is adopted from Ref. [30], where they also derive the ETF energy functional. When we use this energy functional to describe the electron density and couple this to Poisson's equation to complete the self-consistent electrostatic loop, we will refer to it as the ETF method. Below we will minimise the 2D energy functional in Eq. (4.33). For simplicity, we will do the derivation for the 2D case, but it can easily be generalised to 3D. For the simulations in this thesis, we will use the 3D case.

As in the TF case, the mass may indeed have spatial dependence. However, this will not play a role here, and we will thus suppress the dependence for simplicity. We will now introduce the so-called von Weizsäcker field $\psi(\mathbf{r}) \equiv \sqrt{n(\mathbf{r})}$ [26], and write Eq. (4.33) with this, instead of $n(\mathbf{r})$.

$$E_{\text{tot}} = \int \left(\frac{\hbar^2 \pi}{2m} \psi(\mathbf{r})^4 + \lambda_{\text{vW}} \frac{\hbar^2}{8m} \frac{|\boldsymbol{\nabla}(\psi(\mathbf{r})^2)|^2}{\psi(\mathbf{r})^2} + e\psi(\mathbf{r})^2 V_{\text{ext}}(\mathbf{r}) + \frac{1}{2} e^2 \int \frac{\psi(\mathbf{r})^2 \psi(\mathbf{r}')^2}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \right) d\mathbf{r}.$$
$$(4.34)$$

It should be stressed that $\psi(\mathbf{r})$ is *not* the wavefunction of the electron density; it is just a field that squares to the density.

As in the TF case, we will introduce a Lagrangian multiplier μ, but here we will do the variation with respect to $\psi(\mathbf{r})$. We again note that we can rewrite μN by using $\int n(\mathbf{r}) d\mathbf{r} = \int \psi(\mathbf{r})^2 d\mathbf{r} = N$, and thus $\mu\psi(\mathbf{r})^2$ can be added to the integrand. However, the total energy *does* depend on the derivative of the density in this case. This means that we cannot just differentiate with respect to $\psi(\mathbf{r})$, but that we have to calculate the variation as [25]

$$\frac{\delta \mathcal{L}}{\delta \psi(\mathbf{r})} = \frac{\partial(t + v - \mu\psi^2(\mathbf{r}))}{\partial \psi(\mathbf{r})} - \boldsymbol{\nabla} \cdot \frac{\partial(t + v - \mu\psi^2(\mathbf{r}))}{\partial \boldsymbol{\nabla}\psi(\mathbf{r})} = 0. \quad (4.35)$$

here we again used the notation of kinetic and potential energy densities, t and v, including the vW correction in the kinetic part. Doing the differentiation above, we get

$$\frac{\delta \mathcal{L}}{\delta \psi(\mathbf{r})} = \frac{\hbar^2 \pi}{2m} 4\psi(\mathbf{r})^3 - \lambda_{\text{vW}} \frac{\hbar^2}{8m} \frac{\partial}{\partial \psi(\mathbf{r})} \left(\frac{|\boldsymbol{\nabla}(\psi(\mathbf{r})^2)|^2}{\psi(\mathbf{r})^2} \right) + 2eV_{\text{ext}}(\mathbf{r})\psi(\mathbf{r})$$
$$+ \frac{1}{2} e^2 4 \int \frac{\psi^3(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' - 2\mu\psi(\mathbf{r}) - \lambda_{\text{vW}} \frac{\hbar^2}{8m} \boldsymbol{\nabla} \cdot \frac{\partial}{\partial \boldsymbol{\nabla}\psi(\mathbf{r})} \left(\frac{|\boldsymbol{\nabla}(\psi(\mathbf{r})^2)|^2}{\psi(\mathbf{r})^2} \right) = 0.$$
$$(4.36)$$

We will first take a look at the second term and see that this indeed vanishes

$$\lambda_{\text{vW}}\frac{\hbar^2}{8m}\frac{\partial}{\partial\psi(\mathbf{r})}\left(\frac{|\boldsymbol{\nabla}(\psi(\mathbf{r})^2)|^2}{\psi(\mathbf{r})^2}\right) = \lambda_{\text{vW}}\frac{\hbar^2}{8m}\frac{\partial}{\partial\psi(\mathbf{r})}\left(\frac{|2\psi(\mathbf{r})\boldsymbol{\nabla}\psi(\mathbf{r})|^2}{\psi(\mathbf{r})^2}\right)$$
$$= \lambda_{\text{vW}}\frac{\hbar^2}{8m}\frac{\partial}{\partial\psi(\mathbf{r})}\left(4(\boldsymbol{\nabla}\psi(\mathbf{r}))^2\right) = 0. \tag{4.37}$$

We will also simplify the last term of Eq. (4.36). We see that

$$\lambda_{\text{vW}}\frac{\hbar^2}{8m}\boldsymbol{\nabla}\cdot\left(\frac{\partial}{\partial\boldsymbol{\nabla}\psi(\mathbf{r})}\left(\frac{|\boldsymbol{\nabla}(\psi(\mathbf{r})^2)|^2}{\psi(\mathbf{r})^2}\right)\right) = \lambda_{\text{vW}}\frac{\hbar^2}{8m}\boldsymbol{\nabla}\cdot\left(\frac{\partial}{\partial\boldsymbol{\nabla}\psi(\mathbf{r})}\left(4(\boldsymbol{\nabla}\psi(\mathbf{r}))^2\right)\right)$$
$$= \lambda_{\text{vW}}\frac{\hbar^2}{8m}8\boldsymbol{\nabla}\cdot\left(\boldsymbol{\nabla}\psi(\mathbf{r})\right) = \lambda_{\text{vW}}\frac{\hbar^2}{8m}8\boldsymbol{\nabla}^2\psi(\mathbf{r}). \tag{4.38}$$

Again by introducing the electrostatic potential caused by $n(\mathbf{r})$, this time expressed with $\psi(\mathbf{r})$ as

$$u(\mathbf{r}) = e\int\frac{\psi^2(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}', \tag{4.39}$$

we can write Eq. (4.36) as

$$\frac{\delta\mathcal{L}}{\delta\psi(\mathbf{r})} = 2\frac{\hbar^2\pi}{m}\psi(\mathbf{r})^3 + 2V(\mathbf{r})\psi(\mathbf{r}) - 2\mu\psi(\mathbf{r}) - \lambda_{\text{vW}}\frac{\hbar^2}{m}\boldsymbol{\nabla}^2\psi(\mathbf{r}) = 0, \tag{4.40}$$

where we renamed $\mathbf{r}'$ as $\mathbf{r}$ in the $u(\mathbf{r})$ term, and defined again $V(\mathbf{r}) = eV_{\text{ext}}(\mathbf{r}) + eu(\mathbf{r})$. Dividing the equation by 2, and moving around we get the following [26]:

$$\mu\psi(\mathbf{r}) = V_{\text{eff}}(\mathbf{r})\psi(\mathbf{r}) - \lambda_{\text{vW}}\frac{\hbar^2}{2m}\boldsymbol{\nabla}^2\psi(\mathbf{r}), \tag{4.41}$$

where

$$V_{\text{eff}}(\mathbf{r}) = \frac{\hbar^2\pi}{m}\psi(\mathbf{r})^2 + V(\mathbf{r}). \tag{4.42}$$

We see that the Euler-Lagrange equation in (4.41) takes the form of a nonlinear Schrödinger equation [26]. Since the effective potential depends on $\psi(\mathbf{r})$, the nonlinear Schrödinger equation has to be solved self-consistently [29].

The 3D version of the nonlinear Schrödinger equation in (4.41) is [29]

$$\mu\psi(\mathbf{r}) = V_{\text{eff, 3D}}(\mathbf{r})\psi(\mathbf{r}) - \lambda_{\text{vW}}\frac{\hbar^2}{2m}\boldsymbol{\nabla}^2\psi(\mathbf{r}), \tag{4.43}$$

where

$$V_{\text{eff, 3D}}(\mathbf{r}) = \frac{\hbar^2(3\pi^2)^{2/3}}{2m}\psi(\mathbf{r})^{4/3} + V(\mathbf{r}). \tag{4.44}$$

The nonlinear Schrödinger equation in (4.43) is what we will couple to Poisson's equation.

4.4.1 Recovering the TF approximation

Starting from the nonlinear Schrödinger equation in (4.43), we can write it in the following way

$$-\lambda_{\mathrm{vW}}\frac{\hbar^2}{2m}\boldsymbol{\nabla}^2\psi(\mathbf{r}) + \frac{5}{3}C_3\psi(\mathbf{r})^{7/3} + V(\mathbf{r})\psi(\mathbf{r}) = \mu\psi(\mathbf{r}), \tag{4.45}$$

where $C_3 = (3\pi^2)^{2/3}\frac{3\hbar^2}{10m(\mathbf{r})}$. We now wish to investigate what happens if we remove the vW part of the equation, and note that the equation should reduce to the usual TF density, as can be seen in Eq. (4.31). Removing now the vW part, we get

$$\frac{5}{3}C_3\psi(\mathbf{r})^{7/3} + V(\mathbf{r})\psi(\mathbf{r}) = \mu\psi(\mathbf{r}). \tag{4.46}$$

Dividing by $\psi(\mathbf{r})$ we get

$$\frac{5}{3}C_3\psi(\mathbf{r})^{4/3} + V(\mathbf{r}) = \mu. \tag{4.47}$$

Rewriting to the density through the vW field $\psi(\mathbf{r}) = \sqrt{n(\mathbf{r})}$, we get

$$\frac{5}{3}C_3 n(\mathbf{r})^{2/3} + V(\mathbf{r}) = \mu. \tag{4.48}$$

Isolating the charge density $n(\mathbf{r})$ we get

$$\begin{aligned}
n(\mathbf{r}) = \left(\frac{3}{5}C_3^{-1}\right)^{3/2}\left(\mu - V(\mathbf{r})\right)^{3/2} &= \left(\frac{3}{5}\frac{10m}{3\hbar^2(3\pi^2)^{2/3}}\right)^{3/2}\left(\mu - V(\mathbf{r})\right)^{3/2} \\
&= \frac{1}{3\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}\left(\mu - V(\mathbf{r})\right)^{3/2}.
\end{aligned} \tag{4.49}$$

Here we wrote out $C_3 = (3\pi^2)^{2/3}\frac{3\hbar^2}{10m(\mathbf{r})}$. In the above we will also require that $(\mu - V(\mathbf{r})) > 0$, as it is lifted to the power of $3/2$. For the 2D case, we find that it is lifted to the power of $2/2$, and thus one might think that this requirement is not needed in the 2D case. However, as $n(\mathbf{r})$ is the electron density, we will expect this to be larger than zero. Thus one has to physically require that $(\mu - V(\mathbf{r})) > 0$ in both 2D and 3D. We note that Eq. (4.49) is exactly the 3D TF density from Eq. (4.31). We have thus shown that the vW energy functional reduces to the TF density, in the case where the vW term is removed (or stated as the case where $\lambda_{\mathrm{vW}} = 0$).

4.4.2 Value and interpretation of λ_{vW}

So far, we have considered the vW parameter λ_{vW} as a constant, but not discussed its value. Von Weizsäcker's original equation had $\lambda_{\mathrm{vW}} = 1$ [31], however as is shown in Ref. [26], it is actually N-dependant. Historically, the value of λ_{vW} has been extensively discussed in the literature [30]. For instance, Ref. [26] treated λ_{vW} as a fitting parameter, and argued that for a 2D system with N in the range $10^2 - 10^6$, λ_{vW} is in the range of $\sim 0.02 - 0.04$. However, Ref. [26] argued that using a Wigner-Kirkwood (WK)

expansion on physically smooth densities in 2D, all gradient corrections contributing to the kinetic energy functional vanish. For a 3D system Ref. [26] argued that the value of λ_{vW} was closer to ~ 0.25.

The problem of interpreting λ_{vW} in 3D has been clarified by Jones and Young in Ref. [31]. They show that $\lambda_{\mathrm{vW}} = 1$ and $\lambda_{\mathrm{vW}} = 1/9$ are valid in certain regimes. Jones and Young plot $\lambda_{\mathrm{vW}} = 1$ and $\lambda_{\mathrm{vW}} = 1/9$ against exact results for a uniform system of independent fermions, affected by perturbations of different wavelengths. They then conclude that $\lambda_{\mathrm{vW}} = 1$ and $\lambda_{\mathrm{vW}} = 1/9$ are exact in each of their regimes. This is shown in Fig. 4.2 (figure taken from Ref. [31]).
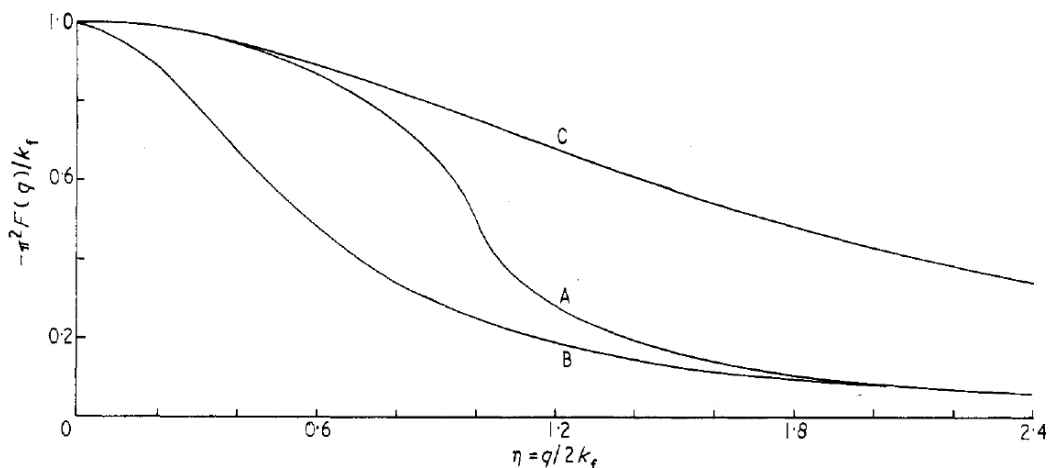

Figure 4.2: Figure taken from Ref. [31]. Response function shown on the y-axis, and relative wavenumber of the perturbation on the x-axis. Curve A denotes the exact results, B denotes $\lambda_{\mathrm{vW}} = 1$, and C denotes $\lambda_{\mathrm{vW}} = 1/9$

In Fig. 4.2 we see how $\lambda_{\mathrm{vW}} = 1$ (B) becomes asymptotically exact for short wavelength perturbations, and $\lambda_{\mathrm{vW}} = 1/9$ (C) becomes asymptotically exact for long wavelength perturbations. For a more detailed description we will refer to Ref. [31].

As it can be extremely difficult to judge exactly what regime an experimental setup belongs to, λ_{vW} can be considered as a fitting parameter. λ_{vW} can be fitted to results of algorithms such as the SP method [28]. However, if one is easily able to run a self-consistent SP algorithm on the entire system, the TF and ETF methods are probably not of any use, as the SP method is generally closer to the exact solution (recalling that the advantage of the TF and ETF methods generally is the speed). Thus, a possible solution is to simplify the experimental geometry and fit the value of λ_{vW} to a self-consistent SP simulation of the simplified geometry. This is a solution used and further discussed later in this thesis.

4.5 Orbital-free model for nanoelectronic devices

In the sections above, we have theoretically used orbital-free theory to express energy functionals for the TF and ETF methods. When applied to real devices, we will assign

physical values to the variables, and the functional we will minimise is thus

$$E[\rho, \phi] = \int \left(C_3 \psi^{10/3}(\mathbf{r}) + \lambda_{\mathrm{vW}} \frac{\hbar^2}{8m(\mathbf{r})} \frac{|\boldsymbol{\nabla}(\psi(\mathbf{r})^2)|^2}{\psi(\mathbf{r})^2} + [-e\phi(\mathbf{r}) - E_c]\psi^2(\mathbf{r}) + \frac{\epsilon |\boldsymbol{\nabla}\phi(\mathbf{r})|^2}{2} \right) d\mathbf{r},$$
(4.50)

where $C_3 = (3\pi^2)^{2/3} \frac{3\hbar^2}{10m(\mathbf{r})}$. The minimisation of this energy functional leads to the non-linear Schrödinger equation. When coupling this to Poisson's equation we get the two equations

$$\begin{cases} -\boldsymbol{\nabla} \cdot (\epsilon \boldsymbol{\nabla}\phi(\mathbf{r})) = -e\psi^2(\mathbf{r}) \\ -\frac{\hbar^2}{2}\lambda_{\mathrm{vW}}\boldsymbol{\nabla} \cdot (m^{-1}(\mathbf{r})\boldsymbol{\nabla}\psi(\mathbf{r})) + \frac{5}{3}C_3\psi^{7/3}(\mathbf{r}) + [-e\varphi(\mathbf{r}) - E_c(\mathbf{r})]\psi(\mathbf{r}) = 0. \end{cases}$$
(4.51)

These are the two equations we will solve when applying the TF and ETF methods to real devices. Here we included the conduction band minimum E_c instead of the more general μ as described above. Note here that we included the vW correction as well. This part is not included in the TF method, i.e. here we interpret the vW coefficient as $\lambda_{\mathrm{vW}} = 0$. The value of λ_{vW} in the ETF method is further described in section 6.4.

4.5.1 Boundary conditions

In this section, we will describe the boundary conditions of the numerical methods.

Boundary conditions for the electric potential

In the electrostatics described above, we have applied a Dirichlet boundary condition of a fixed electric potential whenever describing a gate or wire. A Neumann boundary condition sets the electric field to zero on all semiconductor-vacuum interfaces. These boundary conditions are exact in our case as we consider the electric field explicitly.

Boundary conditions for the pseudo-orbital field ψ in the ETF method

The TF method is entirely described by the electrostatic boundary conditions of the electric field and potential, i.e. Dirichlet of $V = V_{\mathrm{G}}$ on metal-semiconductor interfaces and Neumann of $E = 0$ on all other boundaries. As is further discussed below, the ETF method can have tails out in the regions forbidden by the TF method, and thus the electrostatic boundary conditions do not fully close the problem. This opens the discussion of how to apply correct boundary conditions for ψ in the ETF method. As we expect the dielectric areas to have very low electron density, we expect a Dirichlet boundary condition of $\psi = 0$ to work well at dielectric-semiconductor interfaces. However, we do not expect low electron density in the metals, and thus metal-semiconductor interfaces are more challenging to describe. Below we will discuss the boundary condition for ψ at metal-semiconductor interfaces.

We have tried several different boundary conditions for ψ in the ETF method. To get an intuition of the different conditions, we have made a test case, identical to the one discussed in Ref. [4]. This consists of 100 nm InAs, with a grounded aluminium wire applied at $x = 0$, that has an offset between the aluminium Fermi level and the InAs

conduction band minimum of $0.1\,\mathrm{eV}$. On the other side of the InAs at $x = 100\,\mathrm{nm}$, we apply a Dirichlet boundary condition of $V = -0.3\,\mathrm{V}$. We simulate the electron density in this model using the TF method as well as the ETF method. For the ETF method, we apply a Dirichlet boundary condition of $\psi = 0$ at $x = 100\,\mathrm{nm}$, and try out three different boundary conditions at $x = 0$. The three boundary conditions we tried were a Neumann of $\frac{\partial \psi(0)}{\partial x} = 0$, and two Dirichlet of $\psi(0) = \sqrt{n_{\mathrm{Al}}}$ and $\psi(0) = 0$, where n_{Al} is the electron density of bulk aluminium, which we take to be $1.81 \cdot 10^{29}$ $1/\mathrm{m}^3$ [10]. The results can be seen in Fig. 4.3.
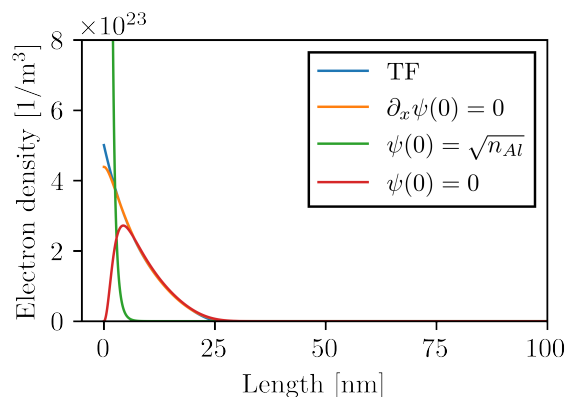

Figure 4.3: Electron density of the simplified Al-InAs model. Gate voltage of $-0.3\,\mathrm{V}$ applied at $x = 100\,\mathrm{nm}$, and assuming $\psi(100\,\mathrm{nm}) = 0$. Calculated for the TF method (blue), and for three different boundary conditions of the ETF method at $x = 0$ (orange, green and red).

Ref. [4] finds reasonably good compliance between the SP method and the TF method, when not too close to the interface. In Fig. 4.3 we see that the two boundary conditions of the ETF method $\frac{\partial \psi(0)}{\partial x} = 0$ and $\psi(0) = 0$ seem to match the TF method nicely when more than ~ 5 nm away from the aluminium. We will thus assume that both of these boundary conditions are reasonably good. However, we see that $\psi(0) = \sqrt{n_{\mathrm{Al}}}$ seems to fail completely as a boundary condition. Thus we must conclude that $\psi(0) = \sqrt{n_{\mathrm{Al}}}$ is a poor choice of boundary condition for the ETF method.

A more thorough investigation of the boundary conditions would be of great interest, as the boundary conditions are crucial to the simulations. However, for now we will settle with the assumption that $\psi(0) = \sqrt{n_{\mathrm{Al}}}$ fails, and both $\frac{\partial \psi(0)}{\partial x} = 0$ and $\psi(0) = 0$ are decent choices for boundary conditions. For the simulations described in section 6, we will use the boundary condition of $\psi(0) = 0$.

Chapter 5

Numerical methods

All numerical work in this thesis is done using COMSOL Multiphysics or Python. COMSOL Multiphysics is a general-purpose simulation software which uses the method of finite elements. It has several built-in methods to solve numerical problems [32]. Naturally, Python does not have an interface and model builder as COMSOL Multiphysics have. Since Python is open-source, it has many packages of algorithms such as NumPy [33] and FEniCS [34], which are the packages used in this thesis. These packages allow the user to easily solve eigenvalue problems and linear problems, which are very relevant for this thesis. In Python, it is essential to understand the numerics, as the problem has to be set up manually. In COMSOL Multiphysics, this is not strictly needed, but when it comes to judging meshes, debugging the solver, and reaching convergence, a numerical understanding and intuition is essential. Therefore we will briefly discuss the basics of numerics needed in this thesis.

For our numerical simulations, we will use two very general methods: The method of finite elements [35], and the method of finite differences. Below we will briefly describe the method of finite differences and show how such a method can be used to solve Poisson's equation as well as the Schrödinger equation, as this will be essential to the work in this thesis. One should note that some of the programs used in this thesis are finite elements based, but we will not go into further detail about how the method of finite elements works. We will instead refer to Ref. [35], and argue that much of the intuition and understanding from the finite difference method can be applied to the finite elements method as well.

The finite difference method is a general way of numerically solving differential equations. All the physical systems we are considering have some continuity, and thus an exact solution would, in principle, require infinitely many calculations, which is impossible [36]. Thus it is essential to discretize the problem, which the method of finite differences describes. Here the system is linearly split into N parts, leaving a system of size L, split into parts of length $L/N = \mathrm{d}x$. This forms a discrete lattice, where each site can be labelled by i, ranging from 0 to N in steps of 1. Note here that we choose a 1D system for simplicity, but the method can be generalized to higher dimensions as well. The method of finite differences approximates the derivative of function $f(x)$ [35],

$$f'(x) = \lim_{\mathrm{d}x \to 0} \frac{f(x + \mathrm{d}x) - f(x)}{\mathrm{d}x}, \tag{5.1}$$

as

$$f'(x) \approx \frac{f(x + \mathrm{d}x) - f(x)}{\mathrm{d}x}, \quad \text{for small } \mathrm{d}x. \tag{5.2}$$

Numerically, this can be done in several different ways. Below three common ways are described [36].

- *Forward finite difference* takes the function value at site $i + 1$ and subtracts the

function value at i, and divides this by $\mathrm{d}x$. Thus you get the average slope to the site after i.

- *Backwards finite difference* takes the function value at site i and subtracts the function value at $i-1$, and divides this by $\mathrm{d}x$. Thus you get the average slope to the site before i.

- *Centered finite difference* takes the function value at site $i+1$ and subtracts the function value at $i-1$, and divides this by $2\mathrm{d}x$. Thus you get the average slope from $i-1$ to $i+1$, which then represents the slope at i.

For the second order derivative, we can use the expressions above to get the *central second finite difference* [36], which is what we will use for the Poisson and Schrödinger equations:

$$\delta^2 f_i = \frac{f(x_{i+1}) - 2f(x_i) + f(x_{i-1})}{\mathrm{d}x^2}. \tag{5.3}$$

These derivatives will be very beneficial when solving the Schrödinger-Poisson problem, as is further described below.

5.1 Finite difference schemes for the Poisson and Schrödinger problems

As we shall see, the Schrödinger-Poisson method is a self-consistent algorithm that can be used to solve the electrostatics of a given system. Generally speaking, the Schrödinger-Poisson method is a mean-field method that couples electron charge to electric potential. The basic assumption of the theory is the Hartree approximation. The Hartree approximatin assumes that the many-body wave function of the system can be expressed as a product of single particle states, i.e. $\Psi(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N) = \psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2)...\psi_N(\mathbf{r}_N)$, which simplifies the problem [10]. Below we will go into detail with first the Poisson part, and then the Schrödinger part, and illustrate how these equations can be solved in a finite difference scheme. We will show this for 1D systems, but it can be generalized to higher dimensions as well.

5.1.1 Poisson's problem

In order to describe the electrostatic environment in the hybrid devices, we will need to be able to calculate the electric potential ϕ induced by a given charge density ρ. This exact problem is governed by Poisson's equation, as described in Eq. (2.1) [4, 16]. Here $\rho[\phi]$ is given as a functional of ϕ [4], but this is suppressed for simplicity in this section. We will thus write Poisson's equation as

$$\boldsymbol{\nabla} \cdot (-\epsilon \boldsymbol{\nabla} \phi) = \rho. \tag{5.4}$$

Here ϵ is the permittivity of the material in which the charge is located. As can be seen in Eq. (5.4), the electron density ρ induces an electric potential ϕ.

In Eq. (5.4) we generally allow the permittivity to be spatially varying, so it appears inside the divergence. For our purpose, we will assume the permittivity to be constant within each material, which allows us to simplify Eq. (5.4). For heterostructures with several different materials, one can simply assume a constant permittivity for each material and then apply Poisson's equation for each material. Assuming a constant permittivity simplifies Eq. (5.4) to

$$- \epsilon \mathbf{\nabla}^2 \phi = \rho. \tag{5.5}$$

The assumption of constant permittivity naturally gives issues at interfaces between two materials with different permittivity. One can here use a shifted lattice for the permittivity at the interface; see for instance Fig. 5.1 for an illustration of this. To discretize this problem, we will take ϕ and ρ to be vectors of length N. Entrance i of ϕ and ρ then corresponds to the electric potential (for ϕ) or charge density (for ρ) at site i. We then have to construct $-\epsilon \mathbf{\nabla}^2$ as a matrix, which can be done using Eq. (5.3). We get that $-\epsilon \mathbf{\nabla}^2$ in Eq. (5.5) can be written as

$$- \epsilon \mathbf{\nabla}^2 = -\epsilon \frac{1}{dx^2} \begin{pmatrix} -2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & -2 \end{pmatrix}. \tag{5.6}$$

As this matrix is applied to heterostructures, one has to also assign the permittivity ϵ to the matrix. One could assign the permittivity of the material at each site, but one would then have an ill-defined permittivity at the interface between two semiconductors. We avoid this by assigning the permittivities to a lattice, which is shifted with $dx/2$ in relation to the original sites. In this way, each site where the electric potential is defined gets assigned the average permittivity of the shifted site before and the shifted site after. This gives a solution for finding the permittivity at an interface between two different materials. This approach is sketched in Fig. 5.1.
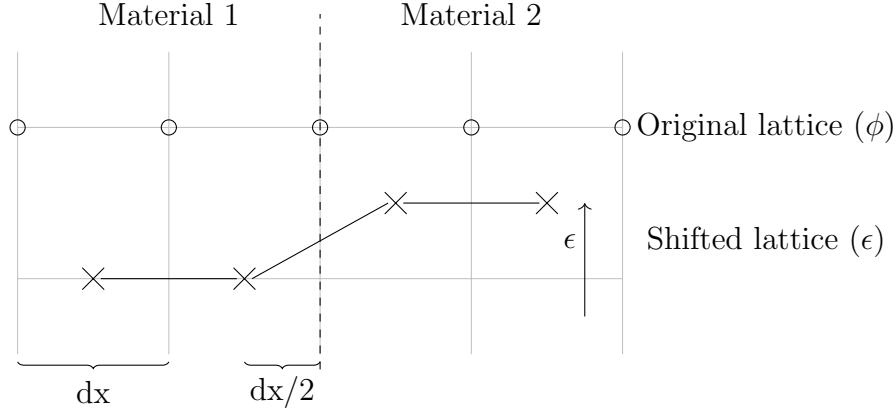
Figure 5.1: Sketch of the two lattices in use for the numerical simulations. The original lattice (circles) are where we differentiate the electric potential. The shifted lattice (crosses) are where we assign the permittivity.

With this discretization, we just need the boundary conditions in order to be able to numerically solve the linear problem $-\epsilon\boldsymbol{\nabla}^2\phi = \rho$. Note that this is exactly of the form $Ax = B$, where A is a matrix, and x and B are vectors. In order to take the boundaries of the system into account, we set the Dirichlet boundary condition in the form of a gate voltage V_G at the first site. This means that the first entrance of ρ should be V_G. On top of that, we will assign the last entrance of ρ to the electric field at the back of the well, E_b. For all simulations discussed in this thesis, we will apply a Neumann boundary condition of $E_b = 0$ to all semiconductors not in contact with a metal. To take these boundary conditions correctly into account, we will also set the lower right entrance of $\boldsymbol{\nabla}^2$ equal to $1/dx$, and the entrance just to the left of it equal to $-1/dx$, as the electric field is minus the differential of the electric potential. Any charge at a given site can then be put into ρ, and any voltage dependence can be put into the $\boldsymbol{\nabla}^2$ matrix. Using this exact approach, we have managed to numerically reproduce the analytical results of the MOS-region in Eq. (3.5).

5.1.2 Schrödinger problem

Considering now the time-independent Schrödinger equation [37]

$$H\Psi = -\frac{\hbar^2}{2m}\boldsymbol{\nabla}^2\Psi + V\Psi = E\Psi, \qquad (5.7)$$

where V is the electric potential of the system, E is the energy, m is the effective mass, Ψ is the complex electron wavefunction, and $|\Psi|^2$ is the normalized electron distribution. The Schrödinger equation in (5.7) is solved for the envelope function of the electron wave functions.

As we are only considering 1D systems, $\boldsymbol{\nabla}^2$ can be constructed the same way as discussed above in section 5.1.1. One should note that since m is the effective mass, one has to change this with the sites, corresponding to the given material of each particular site. Thus a clever trick is to again use the shifted lattice as in Fig. 5.1, and then assign

the mass to this lattice instead of the permittivity. Thus, the effective mass plays the same role in the matrix as the permittivity in the Poisson problem. On top of this, one should subtract the conduction band minimum (see section 2.4.1 for a discussion of the convention) and add the potential multiplied with electron charge, $-e\,V$, to the diagonal of the Hamiltonian. For all the Schrödinger-Poisson simulations discussed in this thesis, we have applied the Schrödinger equation to the semiconducting regions and assumed a Dirichlet boundary condition at the edges of the system, fixing the electron density to 0, i.e. $\psi = 0$. One could argue that the electron density does not necessarily have to be vanishing on all edges, as discussed in Ref. [4]. However, we have assumed that $\psi = 0$ on the edges and will discuss the implications later.

When H is finally constructed, we have the simple eigenvalue problem, $H\Psi = E\Psi$, which can be solved numerically. This gives rise to the electron distribution of the system, $|\Psi|^2$.

5.1.3 Self-consistent Schrödinger-Poisson method

From the above analysis, we see that the Poisson approach takes a charge distribution (ρ) as input and gives the electric potential ϕ as output. The Schrödinger approach takes this electric potential ϕ as input and gives a charge distribution as output. Thus the two approaches form a self-consistent loop. The loop can then be repeated until convergence, as we are considering a time-independent problem. We will name this exact method the Schrödinger-Poisson method (SP method). This is summarised in the flowchart in Fig. 5.2.
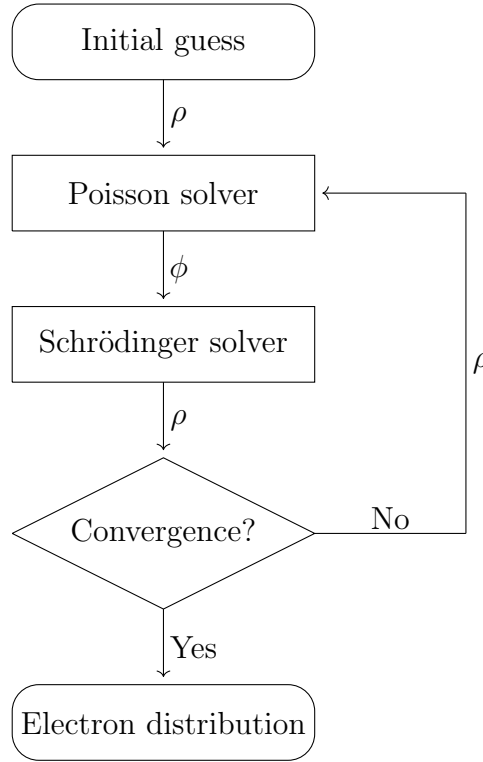
Figure 5.2: Sketch of the Schrödinger-Poisson method (SP method). Starting from an initial guess, the electron distribution of the system is numerically calculated up to the desired precision.

Whether the algorithm converges can be determined from a comparison between the given electron distribution, ρ_i, and the previous, ρ_{i-1}, where i is the number of loops in the calculation. If the relative difference between the charge distribution and the charge distribution from the previous loop is below some threshold value t, one can define this as convergence. Thus we define a criterion for convergence as

$$\left| \frac{\rho_i - \rho_{i-1}}{\rho_{i-1}} \right| < t. \tag{5.8}$$

The relative threshold can then be set to the value of the desired precision. For many of the simulations in this thesis, a threshold of 10^{-5} is used.

If the two electron densities do not fulfil Eq. (5.8), one can discard the old electron density ρ_{i-1} and use the new electron density ρ_i in the next loop. This is what is shown in the flow chart in Fig. 5.2. However, completely discarding the old solution can lead to convergence issues. One way to solve this is to use a mixing scheme, where ρ_{i-1} and ρ_i are mixed before the mixed electron density ρ_i^{mix} is used in the Poisson approach again. Thus the mixed electron density is defined as

$$\rho_i^{\mathrm{mix}} = (1 - \lambda)\rho_{i-1} + \lambda\rho_i. \tag{5.9}$$

Here $\lambda = 1$ corresponds to just taking the new value ρ_i and discarding the old one ρ_{i-1}, as is sketched in Fig. 5.2, and $\lambda = 0$ corresponds to never using a new value, i.e. the

iterations will never leave the initial guess. For most purposes, $\lambda \approx 0.1$ gives a robust convergence [10], but as each iteration of the SP method might be very computationally expensive, one often want a larger λ.

This mixed scheme can also be implemented for the TF and ETF methods, as it generally helps the algorithm converge. The requirement for convergence shown in Eq. (5.8) is also implemented for the TF and ETF methods.

For the initial guess in the SP method, one could for instance use a few iterations of the TF method to get a rough estimate of the electron distribution, or simply use an arbitrary charge distribution such as any of the charge distributions in Fig. 3.3.

5.1.4   Schrödinger-Poisson scheme for translationally invariant systems

The diagonalization needed to solve the Schrödinger equation makes the method very computationally costly. For a 3-dimensional real system, this task may be so large that it is effectively impossible to do it in a reasonable time scale. This lead to our motivation for investigating orbital-free methods. For many of the real systems discussed in this thesis, it is possible to split the systems into simplified regimes. This is, for instance, the case in the MOS-region and the wire-region, as is shown in Fig. 3.1. These two particular regions are generally more simple than the real devices, as they have translational symmetry in the x- and y-directions, and thus only vary in the z-direction, i.e. the direction going through the semiconducting stack. This translational symmetry can be effectively exploited in the calculations to decrease computation time drastically. For the SP method, this can be done by solving in the z-direction only, and then assuming that the solution does not change in the xy-plane.

This approach differs slightly from the algorithm sketched in figure 5.2. Since we only solve in one direction, we will not directly calculate the electron density from the Schrödinger equation. We will instead get the eigenvalues E_n and eigenmodes U_n, and from these calculate the electron density. The eigenmodes give rise to the electron density

$$\rho = -e \sum_n U_n^2 n_{2D} E_n \theta(E_n), \quad n_{2D} = \frac{m_{\text{eff}}}{\pi \hbar^2}. \tag{5.10}$$

Here n_{2D} is the electron density of a 2D gas, as is described in equation (4.30), and $\theta(E_n)$ is the Heaviside step function for E_n. For systems with translational invariance, we will thus have the extra step of calculating the electron density from the eigenmodes, as shown in equation (5.10).

Generally speaking, if we assume N discretization points in all three directions, the problem is of size $N \times N \times N$, and thus using the translational invariance brings the problem to size N. This is one of the methods that we will use extensively to bring down computation time, for instance in section 6.3 and 6.4.

5.1.5 Boundary Conditions of the wave function in the SP method

The SP method described above is made with the Schrödinger equation defined only in the semiconductor, i.e. neglecting states in the metal. Thus we have assigned the Dirichlet boundary condition of $\psi = 0$ on the semiconducting edges. The metal serves then to fix the metal-semiconductor interface at some voltage V_G, i.e. the bands are only filled with electrons corresponding to an energy of $\phi_\mathrm{w} + V_\mathrm{G}$ on this interface. This is naturally a Dirichlet boundary condition in Poisson's equation, i.e. $V = V_\mathrm{G}$. For all semiconducting boundaries not in contact with metal, the Neumann boundary condition of $E = 0$ is applied.

The SP method applied below use this exact same setup. However, as is briefly mentioned above, $\psi = 0$ at the metal-semiconductor interface might not be a good boundary condition. Actually the bulk electron density of aluminium is $\sim 1.81 \cdot 10^{29}$ $1/\mathrm{m}^3$, corresponding to a Fermi energy of ~ 11.7 eV [4, 10]. Thus one could argue, that ψ on the metal-semiconductor interface should correspond to the bulk electron density of aluminium, to accommodate a continuous change in electron density across the interface. This exact difference is discussed in Ref. [4, 10], and turns out to have a very large difference close to the metal-semiconductor interface, but the two methods coincides when just (roughly, see Ref. [4, 10] for details) ~ 5 nm from the interface. Thus for the systems where we have a barrier between the metal and the well, we suspect this difference in boundary conditions to play a relatively small role. On top of this, one of the basic assumptions is to neglect surface states. This definitely plays a large role in this regard, and as we do not fully understand all the surface physics, we will for the SP method use this simplified boundary condition of $\psi = 0$ on the edges.

Another approach for understanding the $\psi = 0$ boundary condition could be to visualise an infinitesimally thin, perfect insulator at the boundary between metal and semiconductor. In this way, the electric potential from the metal would still give a Dirichlet boundary condition for Poisson's equation but apply $\psi = 0$ for the interface when the Schrödinger equation is solved in the semiconductor.

So far, we have only applied the SP method to the semiconducting areas. One could also apply the SP method in the metal, and thus self-consistently solve the problem for the entire device domain. However, this is a more advanced approach. For now we will stick to the method of only applying the SP method to the semiconducting regions and refer the interested reader to Ref. [4, 10].

5.2 Numerical use of the extended TF functional

Given the complexity of the device geometry, we will model the devices using finite element modelling (FEM). As a prerequisite of FEM, the differential problem has to be written in its weak formulation. This is done automatically for the Poisson problem in COMSOL Multiphysics, but for the nonlinear Schrödinger equation in (4.43), the weak form has to be analytically derived. This is naturally because the Poisson problem is

a very widespread problem, and thus built-in functions are handling it, whereas the nonlinear Schrödinger equation is an uncommon problem. Below we will describe how to formulate the nonlinear Schrödinger equation from the ETF method in its weak formulation, such that it can be minimised with respect to $\psi(\mathbf{r})$.

5.2.1 Weak formulation of the extended TF functional

The method of weak formulation is a widely used tool for solving partial differential equations (PDEs) [35]. The core concept is to turn a PDE into a variational problem by multiplying with a so-called test function v, integrating over the entire system Ω, and then using integration by parts to reduce possible second-order derivatives [35]. The unknown function of the PDE, called the trial function, can then be calculated by means of variation [35]. Below we will apply this method to the present case of the ETF energy functional.

We will start by considering the nonlinear Schrödinger equation from the ETF functional in Eq. (4.43). Note here that we do take the spatial dependence of the mass into account and thus get that:

$$-\lambda_{\mathrm{vW}}\frac{\hbar^2}{2}\boldsymbol{\nabla}\cdot\left(m^{-1}(\mathbf{r})\boldsymbol{\nabla}\psi(\mathbf{r})\right) + \frac{5}{3}C_3\psi(\mathbf{r})^{7/3} + V(\mathbf{r})\psi(\mathbf{r}) = \mu\psi(\mathbf{r}).\qquad(5.11)$$

Introducing now the differentiable test function $v(\mathbf{r})$, we can write the weak form of the nonlinear Schrödinger equation as

$$-\int_\Omega \lambda_{\mathrm{vW}}\frac{\hbar^2}{2}\boldsymbol{\nabla}\cdot\left(m^{-1}(\mathbf{r})\boldsymbol{\nabla}\psi(\mathbf{r})\right)v(\mathbf{r})d\mathbf{r} + \frac{5}{3}C_3\int_\Omega \psi(\mathbf{r})^{7/3}v(\mathbf{r})d\mathbf{r} + \int_\Omega (V(\mathbf{r})-\mu)\psi(\mathbf{r})v(\mathbf{r})d\mathbf{r} = 0.$$
$$(5.12)$$

Here we integrate over the entire system volume Ω and require $v(\mathbf{r})$ to be 0 at the boundary of the system $\partial\Omega$, i.e.

$$v(\mathbf{r}) = 0, \quad \forall\, \mathbf{r} \in \partial\Omega.\qquad(5.13)$$

Using partial integration on the first term of Eq. (5.12) we get

$$-\int_\Omega \lambda_{\mathrm{vW}}\frac{\hbar^2}{2}\boldsymbol{\nabla}\cdot\left(m^{-1}(\mathbf{r})\boldsymbol{\nabla}\psi(\mathbf{r})\right)v(\mathbf{r})d\mathbf{r} = -\int_{\partial\Omega} \lambda_{\mathrm{vW}}\frac{\hbar^2}{2}\left(m^{-1}(\mathbf{r})\boldsymbol{\nabla}\psi(\mathbf{r})\right)v(\mathbf{r})d\mathbf{r}$$
$$+\int_\Omega \lambda_{\mathrm{vW}}\frac{\hbar^2}{2}m^{-1}(\mathbf{r})\boldsymbol{\nabla}\psi(\mathbf{r})\cdot\boldsymbol{\nabla}v(\mathbf{r})d\mathbf{r}.$$
$$(5.14)$$

However, as we define $v(\mathbf{r})$ to be zero at the boundary of Ω, the first term on the right side must be equal to zero. Thus we get

$$-\int_\Omega \lambda_{\mathrm{vW}}\frac{\hbar^2}{2}\boldsymbol{\nabla}\cdot\left(m^{-1}(\mathbf{r})\boldsymbol{\nabla}\psi(\mathbf{r})\right)v(\mathbf{r})d\mathbf{r} = \int_\Omega \lambda_{\mathrm{vW}}\frac{\hbar^2}{2}m^{-1}(\mathbf{r})\boldsymbol{\nabla}\psi(\mathbf{r})\cdot\boldsymbol{\nabla}v(\mathbf{r})d\mathbf{r}.\quad(5.15)$$

The weak form in Eq. (5.12) can therefore be rewritten as

$$\int_\Omega \lambda_{\mathrm{vW}}\frac{\hbar^2}{2}m^{-1}(\mathbf{r})\boldsymbol{\nabla}\psi(\mathbf{r})\cdot\boldsymbol{\nabla}v(\mathbf{r})d\mathbf{r} + \frac{5}{3}C_3\int_\Omega \psi(\mathbf{r})^{7/3}v(\mathbf{r})d\mathbf{r} + \int_\Omega (V(\mathbf{r})-\mu)\psi(\mathbf{r})v(\mathbf{r})d\mathbf{r} = 0.$$
$$(5.16)$$

This weak form can be used for numerical calculations minimising the energy functional. One can find several algorithms that minimise functionals on this form. Thus given an input voltage V, we can use this approach to get an output charge density, $en(\mathbf{r}) = \rho(\mathbf{r})$ (remember that $\psi(\mathbf{r})^2 = n(\mathbf{r})$). With this in mind, one can say that the ETF functional approach serves the same purpose as the Schrödinger approach in the algorithm outlined in Fig. 5.2. Therefore, we can build the algorithm, summarised in the flowchart in Fig. 5.3.
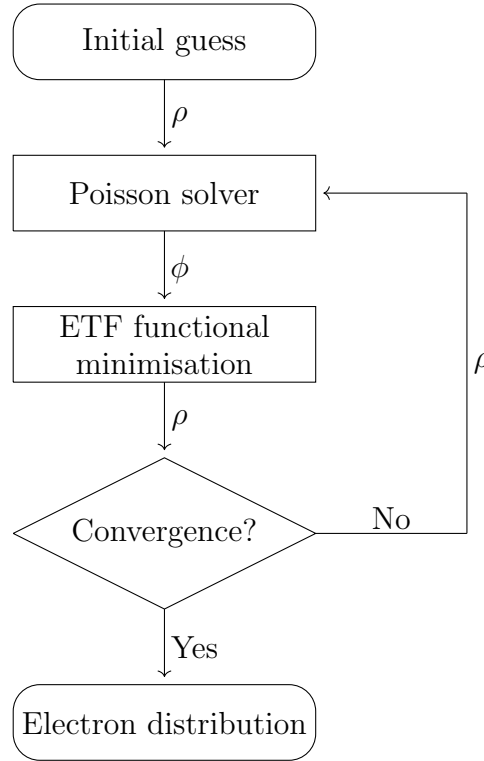
Figure 5.3: Sketch of the extended Thomas-Fermi method (ETF method). Starting from an initial guess, the electron distribution of the system is numerically calculated.

The flow chart in Fig. 5.3 is shown for a segregated approach. This could equally well be solved with a fully coupled approach, where Poisson's equation and the non-linear Schrödinger equation is solved simultaneously. The clever thing about the ETF algorithm in Fig. 5.3, is that it is generally *much* quicker than the SP method in Fig. 5.2, as the diagonalization needed in the Schrödinger part is more computationally expensive than the functional minimisation in the ETF part. On top of this, the vW correction fixes some of the unphysical behaviour of the TF method [4, 26, 29].

5.3   Uncertainty of physical parameters

Above, it is briefly mentioned that the surface states of the interfaces are not fully understood. This definitely plays a role in the experiments, and therefore it would be beneficial to have further knowledge of the behaviour of these surface states. The

distribution of surface states is object to active experimental research. There are several different models for surface states, and in section 6.7.2 we will briefly discuss the simplest one, which is a linear density of states. However, the surface states are not the only thing we do not fully understand. It is believed that the oxide contains charge density, which we do not know the behaviour of. If there is a constant oxide charge in the experiments, which is not taken into account in the simulations, there will be an offset between the experimental gate voltages and the simulated gate voltages. This would be very useful and interesting to investigate further. In some experiments, hysteresis as a function of time and temperature might occur. If a device is cooled down, reheated, and cooled down again, it does not necessarily have the exact same features as before the reheating. It would be useful to gain further insight into this kind of behaviour.

All these topics are difficult to take into account, and thus we have simplified the simulations in this thesis (except for the ones in section 6.7.2 and 6.7.3) by neglecting them.

Chapter 6

Results

Below are the numerical results of the device simulations and a detailed discussion of the implementation. The theory behind this is covered in the sections above.

6.1   TF based simulations

For a first device simulation, the applied gate voltage of the device can be simulated, and the corresponding electron density can be calculated with the TF method, as described in Eq. (4.49). This is indeed a very simplified approach, but since it is very easily implemented and computationally inexpensive, it is a good first hint of what sort of electron distribution to expect. As discussed in the sections above, it can be argued that far from the edges, where the electron density is assumed slowly varying, the TF method alone might be quite a good approximation. Therefore, this method is still widely used and is also used as the base for several of the simulations made in this thesis.

To show how the TF method is practically implemented, we will briefly review how the simulation of a particular device is set up. We will consider the device shown in Fig. 6.1. This device consists of a long aluminium wire (blue in the figure), with so-called plunger gates (yellow in the figure) on top. Along with this, we have 5 so-called cutter gates (orange in the figure) coming in from the right. The plunger gates serve the purpose of depleting around the aluminium wire, thus creating a proximitized electron wire below the aluminium wire. The cutter gates serve the purpose of tuning tunnel barriers between the wire and the probes. The probes are formed by a superconducting aluminium layer below the cutters (blue in the figure). One of the cutters and the underlying aluminium can be seen on the zoom in in Fig. 6.1.

The material stack of this particular device can be seen in Fig. 6.2. Starting from the bottom, we have a 25 nm layer of $In_{0.82}Al_{0.18}As$, working as so-called sub barrier. On top of this we have a 4 nm lower barrier, consisting of $In_{0.75}Ga_{0.25}As$, followed by a 7 nm well consisting of InAs. This is topped off by a 10 nm upper barrier consisting of $In_{0.75}Ga_{0.25}As$. These four layers are what we are considering as the semiconducting stack. On top of the semiconducting stack, all the aluminium is deposited. Thus the aluminium wire and the aluminium below the cutter are in the same layer. On top of this, we have a 15 nm layer of dielectric (HfO_2). The cutter gates (20 nm of Ti/Au) are then deposited on top of this, followed by another dielectric layer. On top of this second layer of dielectric, we have the plunger gates (20 nm of Ti/Au). Thus the cutters and the plungers are shifted by a dielectric layer, as shown in Fig. 6.2.
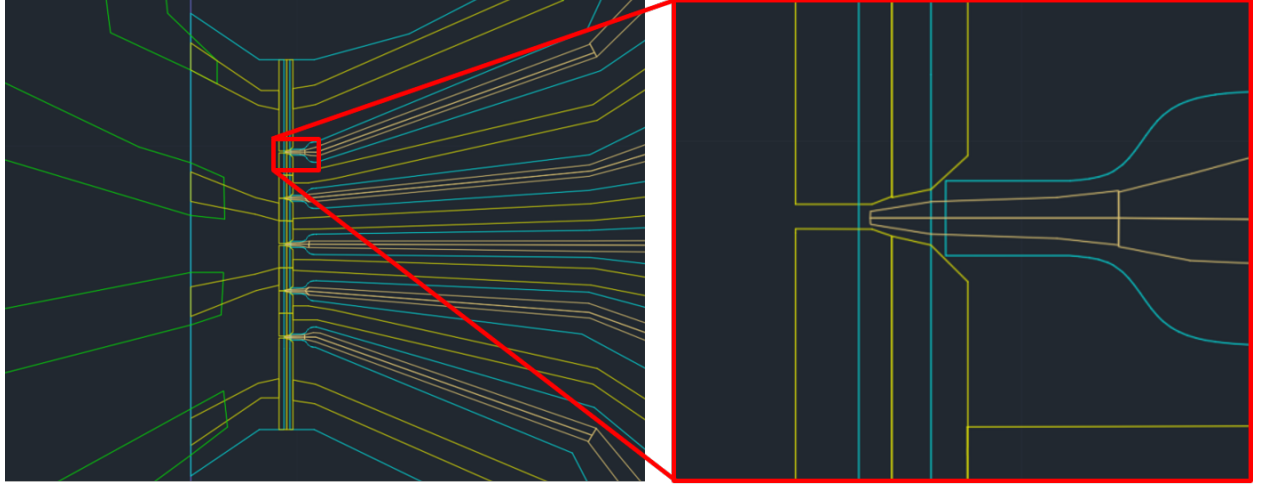
Figure 6.1: CAD drawing of the device, the zoom includes a close-up of one of the cutter gates. Designed and fabricated by Andreas Pöschl et. al.
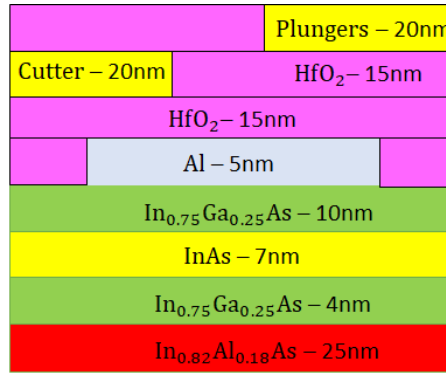

Figure 6.2: Sketch of the stack of the device in Fig. 6.1.

For the TF method, we will apply a Dirichlet boundary condition on the gates (cutters and plungers), in the form of a constant applied electric potential. All of the aluminium will also be applied a Dirichlet boundary condition, but it will be fixed to 0 V, i.e. the aluminium is grounded in these simulations. We should note that the device is so cold that the aluminium is superconducting in the real experiments. Thus we are neglecting superconductivity when using the TF method. For every exterior boundary of the device, a Neumann boundary condition is assumed, electrostatically corresponding to an electric field of zero.

The device in Fig. 6.1 consists of 5 plunger gates, which are generally applied negative voltages, such that they deplete the area around the wire below. For a gate voltage of around -3 V, we expect the semiconductor to be in a topological phase, and for a gate voltage of around -6 V we expect the semiconductor to be in a trivial phase. We will run the simulations with the two outmost plunger gates, i.e. the ones in the bottom and top in Fig. 6.1, at -6 V, and the three inner ones at -3 V. This is such that we get an isolated topological 2DEG wire.

The TF method is applied to the semiconducting stack. Due to the applied potential from the gates (and the grounded aluminium), the semiconducting stack will have some spatially varying electric potential. The electron density can be computed from the effective mass and band offset of each material in the stack through the TF electron density.

This approach is applied to the device in Fig. 6.1. In Fig. 6.3 we see a contour plot of the electric potential (a) and the electron density (b). This serves primarily the purpose of getting an overview of the electrostatics of the device. In Fig. 6.3 (a), we see how the electric potential is relatively large on the wire and the cutter probes, and relatively small in the outer regions of the device. The large electric potential attracts electrons, as shown in 6.3 (b). Here we see the electrons confined to a small area under the wire, with side-probes coming in from the right.

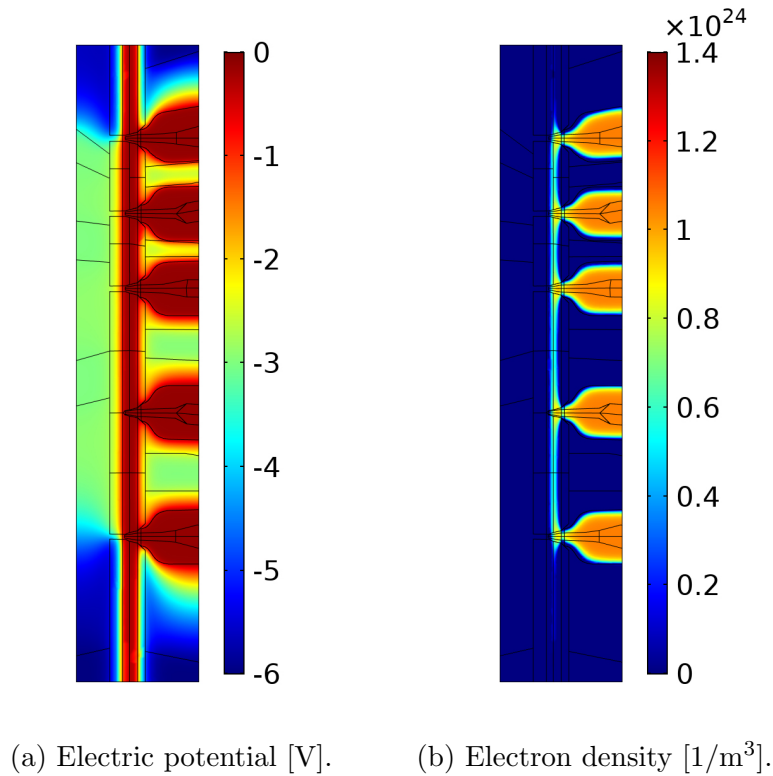(a) Electric potential [V].   (b) Electron density $[1/m^3]$.

Figure 6.3: Plots of the electric potential (a) and the electron density (b), for the entire device. Outer plungers at -6V, inner ones at -3V, and all cutters at -0.1V. Plot made as a slice cut through the middle of the well, using the TF method.

In Fig. 6.4 we see a zoom in of the electron density on the middle sideprobe. Here one can clearly see how the sideprobe probes out the density in the wire. The shape of the potential around the probe region is crucial for the experiment since it is essential for forming a good tunnel barrier. For instance, one could conclude from Fig. 6.4 that a larger tunnel barrier would be needed to form a QPC (since the pinch-off of electron density from the wire to the sideprobe is relatively small in Fig. 6.4).
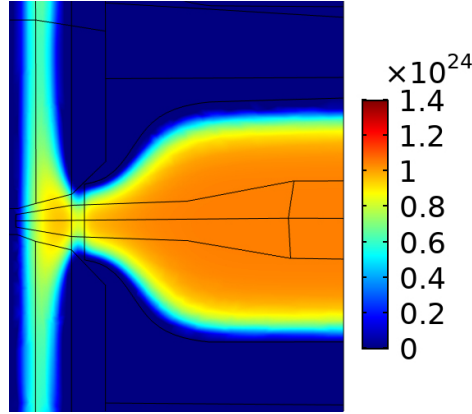
Figure 6.4: Zoom in of the electron density [1/m^3] of the middle cutter from Fig. 6.3 (b). Calculated using the TF method.

These contour plots give a great overview of the electrostatics, but as we want to inspect the results even further, we want to make slice cuts through the device, such that we can get line plots instead of contour plots. In Fig. 6.5 we have cut the wire in the middle of the well, plotting the voltage along this axis. Here we see how the voltage peaks at each sideprobe.
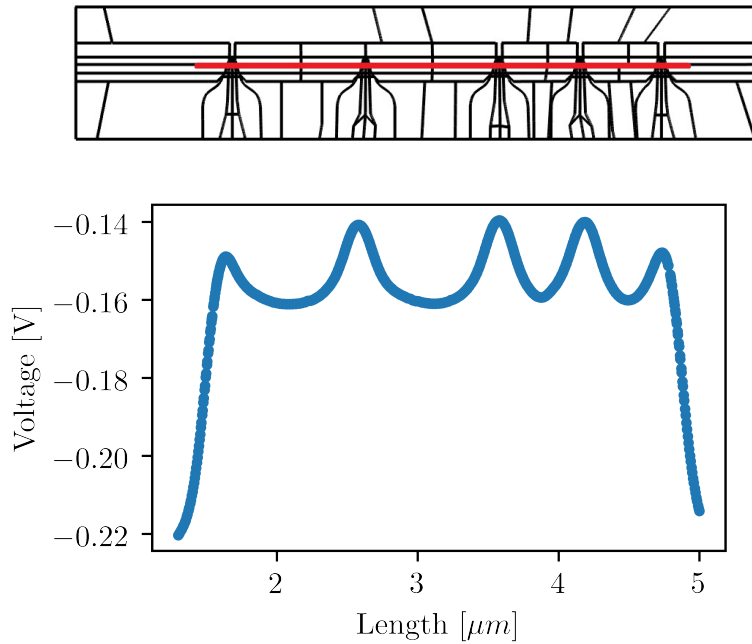

Figure 6.5: Voltage as a function of length along the red line, as shown in the figure above. Calculated using the TF method. We see how each cutter correspond to a peak in voltage. The red line is in the middle of the well.

In Fig. 6.6 we have made a cut through a cutter, plotting again the voltage along this axis. As can be seen, there is a local minimum of the voltage. This local minimum

in the direction of the sideprobe forms a saddle point with the local maximum in the direction of the wire from Fig. 6.5.
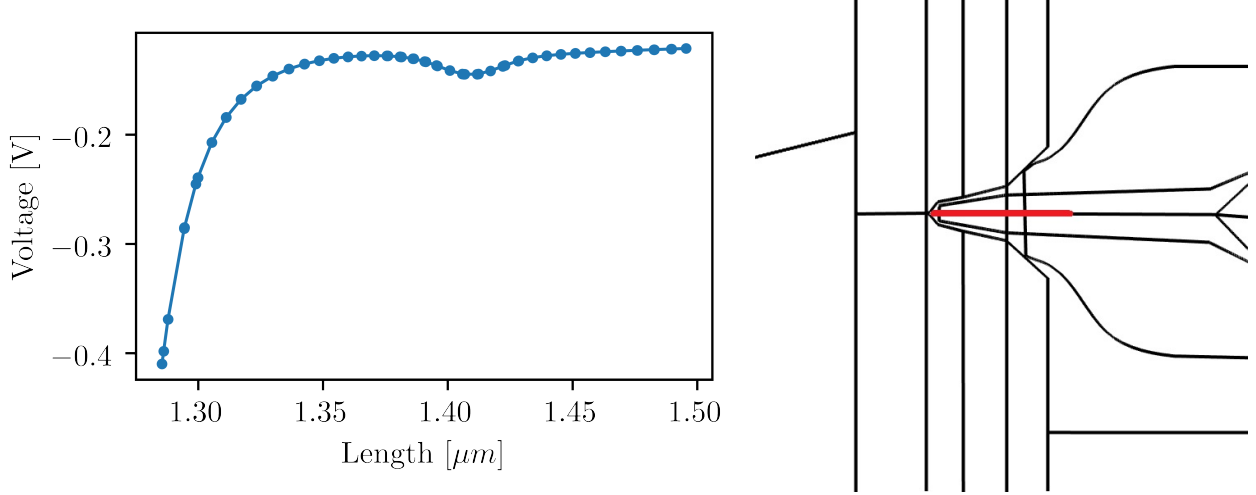

Figure 6.6: Voltage as a function of length along the red line, as shown in the figure to the right. Calculated using the TF method. We see how there is a local minimum in the middle of the plot. The red line is in the middle of the well.

The line plots in Fig. 6.5 and Fig. 6.6 show how the sideprobes disturb the electric potential in the well, and thus how it affects the electron density. In Fig. 6.5 we see how the outer plungers cause a low voltage in the well and how the inner plungers increase the voltage. This is what causes the transition between topological and trivial phases.

6.1.1   Material properties

Throughout this thesis, several semiconducting stacks have been simulated, and thus many different material properties have been used. These are listed here and briefly explained below.

For the semiconductors, we used the values as can be seen in Tab. 6.1.

Property	AlAs	GaAs	InAs
ϵ/ϵ_0	10.06	15.70	15.15
$m_{\text{eff}}/m_{\text{e}}$	0.150	0.067	0.026
$E_g[\text{eV}]$	3.099	0.812	0.417
$E_v[\text{eV}]$	-1.330	-0.800	-0.590

Table 6.1:   Relevant physical properties of the semiconductors used in this thesis. Material properties collected from several references, see Ref. [38, 39, 40].

However, we are using several different semiconducting mixed alloys. These can be found, along with the oxide used in the simulations, in Tab. 6.2. Some of these values are calculated from interpolation of the values in Tab. 6.1.

Property	$In_{0.75}Ga_{0.25}As$	$In_{0.85}Al_{0.15}As$	$In_{0.82}Al_{0.18}As$	$In_{0.81}Al_{0.19}As$	HfO_2
ϵ/ϵ_0	14.76	14.39	14.23	14.18	25.00
$m_{\mathrm{eff}}/m_{\mathrm{e}}$	0.035	0.038	0.041	0.042	-
E_g[eV]	0.603	0.730	0.796	0.819	-
E_v[eV]	-0.571	-0.619	-0.629	-0.632	-
ϕ_{w}[eV]	0.0577	0.0325	-	-	-

Table 6.2: Relevant physical properties of the materials used in the simulations in this thesis. Material properties collected from several references, see Ref. [38, 39, 40].

In Tab. 6.1 and 6.2, ϵ is the permittivity, ϵ_0 is the vacuum permittivity, m_{eff} is the effective electron mass, and m_{e} is the electron mass. E_g is the bandgap, E_v is the valence band maximum, i.e. $E_g = E_c - E_v$, where E_c is the conduction band minimum. ϕ_{w} in Tab. 6.2 is with respect to the Fermi level in aluminium. For instance, we should interpret $\phi_{\mathrm{w}} = 0.0325$ eV for $In_{0.85}Al_{0.15}As$ such that the conduction band minimum of $In_{0.85}Al_{0.15}As$ is $\phi_{\mathrm{w}} = 0.0325$ eV below the aluminium Fermi level. See for instance Fig. 2.2 for a simplified schematic of how the bands are considered in these simulations.

6.2   Weakness of the TF method

We have previously argued that the TF method gives rise to an unphysical electron density, as it predicts very sharp cutoffs. Indeed this is partly caused by the modelling of the changing bandgaps and electron masses as an abrupt change, going from one material to another. The conduction band minimum as a function of length going through the semiconducting region, can be seen in the orange graph in Fig. 6.7. On this plot, one can also see the conduction band minimum minus the electric potential energy $E_c - \phi$ (blue). The potential used here is calculated using the TF method in a cut through the semiconducting stack with a grounded aluminium wire on top, i.e. it corresponds to the simplified wire model, as discussed in section 3.2.
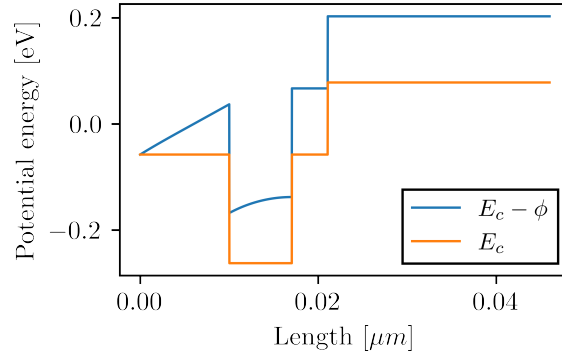
Figure 6.7: Conduction band minimum (orange) and conduction band minimum minus the electric potential energy (blue) of the semiconducting stack in Fig. 6.2. Plotted as a function of distance from the aluminium wire. Simulated as the wire model, i.e. we assume a grounded aluminium wire to the left of the plot. Potential calculated with the TF method.

This abrupt change in bands results in the electron density, as can be seen in Fig. 6.8,
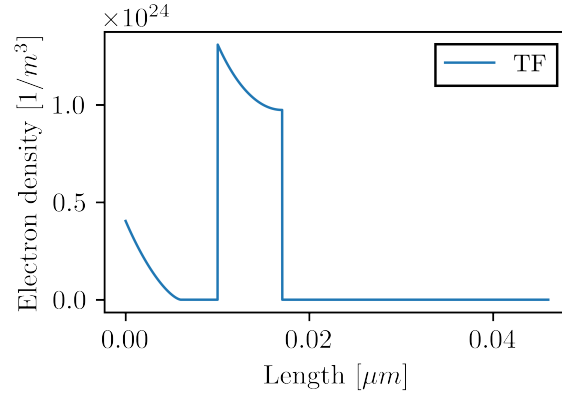

Figure 6.8: Electron density as a function of distance from the wire, calculated from the bands seen in Fig. 6.7. Electron density calculated with the TF method.

As can be very clearly seen in Fig. 6.8, the electrons are quite closely confined to the well. However, we see an extremely sharp cutoff of electron density, which illustrates one of the major issues with the TF method.

6.2.1 Slowly varying electron density

As is briefly described in section 4.3, the TF method assumes a slowly varying electron density. Mathematically this can be expressed as [41, 42]

$$L_{\text{dens}} \equiv \frac{|\boldsymbol{\nabla} n(\mathbf{r})|}{n(\mathbf{r}) k_{\text{F}}(\mathbf{r})} \ll 1, \quad \text{where} \quad k_{\text{F}}(\mathbf{r}) = (3\pi^2 n(\mathbf{r}))^{1/3}. \tag{6.1}$$

Plotting the value of L_{dens} using the electron density from Fig. 6.4, we get an overview of where the electron density is slowly varying, and where it is not. This can be seen in Fig. 6.9.
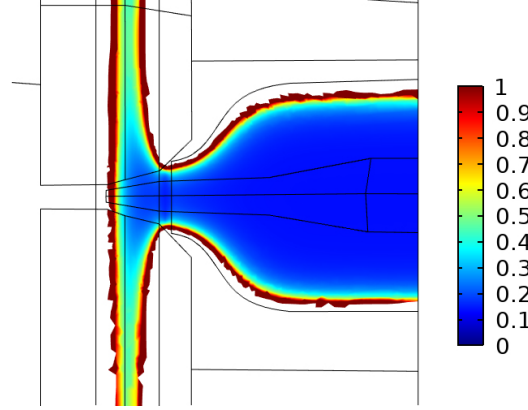

Figure 6.9: Contour plot of L_{dens} [unitless] for the electron density shown in Fig. 6.4. Calculated using the TF method. Color scale of L_{dens} cut at a maximum of 1. Cut of the device is made through the middle of the InAs well.

In Fig. 6.9 we see how L_{dens} is defined everywhere except for the area where the electrons are depleted. As can be clearly seen in Fig. 6.9, there are several areas where the electron density is not slowly varying. The slowest varying areas has $L_{\mathrm{dens}} \approx 0.2$. In particular, the transition between depleted and non-depleted areas naturally have a very rapidly changing electron density, which again underlines the steep cutoff behaviour in the TF predicted electron density.

To further investigate this, we have applied the same method to another kind of device, the so-called dot-probe device. This device can be seen in Fig. 6.10.
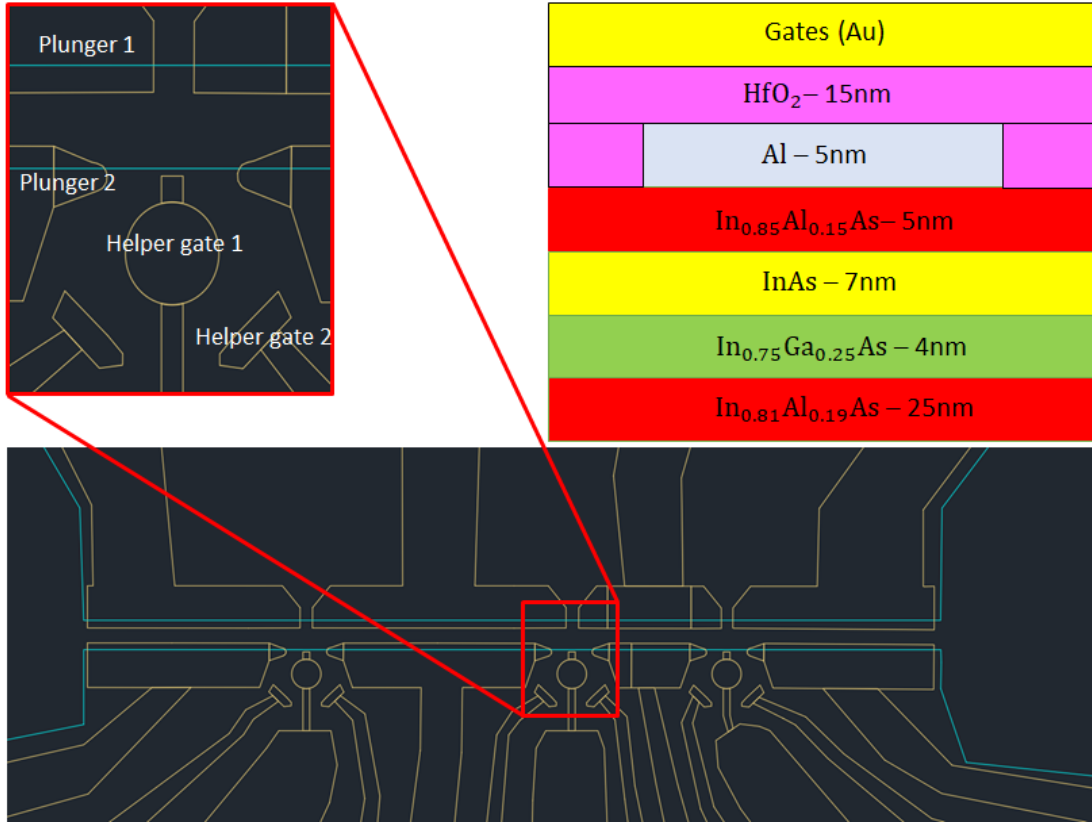
Figure 6.10: CAD-drawing of the dot-probe device, along with the corresponding material stack. Designed and fabricated by Alisa Danilenko et. al.

This device aims to form a quantum dot next to the wire by attracting electrons with helper gate 1. Helper gate 2 then confines the electrons to the area below helper gate 1, and also creates a pinch-off to the electron density. Both plunger 1 and plunger 2 deplete the area next to the wire, and plunger 2 also has the purpose of creating a pinch-off to the quantum dot. A successful device will thus have a clearly defined quantum dot below helper gate 1 and two QPCs made from the pinch-off of electron density (one QPC formed by plunger 2, and one from helper gate 2). The gates are named as seen in Fig. 6.10, and the aluminium is the wire marked as blue in the figure.

We have applied the TF method to this device as well, using the following voltages: Plunger 1 at -4 V, plunger 2 at -2 V, helper gate 1 at 0 V, helper gate 2 at -0.5 V, and the aluminium wire grounded. This gives an electron density along with a plot of L_{dens} as shown in Fig. 6.11.

(a) Electron density $[1/\text{m}^3]$.

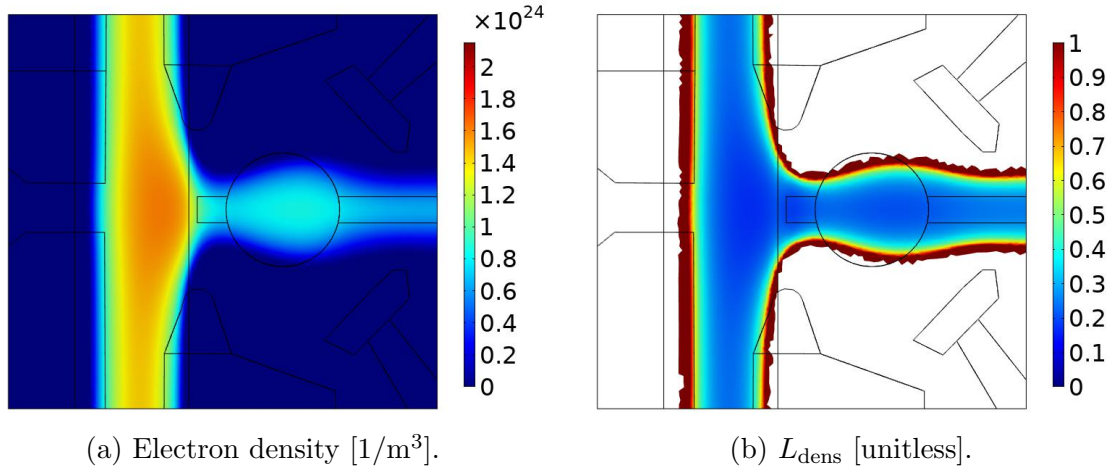(b) L_{dens} [unitless].

Figure 6.11: Plots of the electron density (a) and L_{dens} (b) for the dot-probe device. Calculated using the TF method. Color scale of L_{dens} cut at a maximum of 1. Both cuts are made in the middle of the InAs well.

In Fig. 6.11 (a), we see how the plungers closely confine the electrons to the wire. We also see how helper gate 1 attracts electrons and how helper gate 2 confines them. However, a quantum dot is not entirely formed in Fig. 6.11 (a). In order to form a quantum dot, one needs a lower voltage of plunger 2 and helper gate 2, as this would pinch off the electron density.

Looking at L_{dens} in Fig. 6.11 (b), we see again how it is not $\ll 1$. Thus in these devices, the TF method seems to predict a rather rapidly varying electron density.

6.3   TF method as an initial guess for the SP method

As discussed in section 6.2, the TF method has some weaknesses and limitations. However, the power of the TF method is in the speed and not in the predictive power, at least when compared to the SP method. Therefore, the TF method could be used as a preliminary step to the SP method, such that the initial guess of the SP method is optimized with the TF method. In particular, we will consider the case where the TF method first is used until convergence and then the Schrödinger equation is solved once using the electric potential from the TF method. Therefore, the TF method works as an initial guess for a single diagonalization of the Schrödinger equation, and thus greater speed than a regular full SP method is obtained. We will call this method "TF + 1xS", as we first use the TF method, and then solve the Schrödinger equation once. In Fig. 6.12 we compare this method to a full SP method as well as a full TF method, using the wire model with the dot-probe stack from Fig. 6.10.
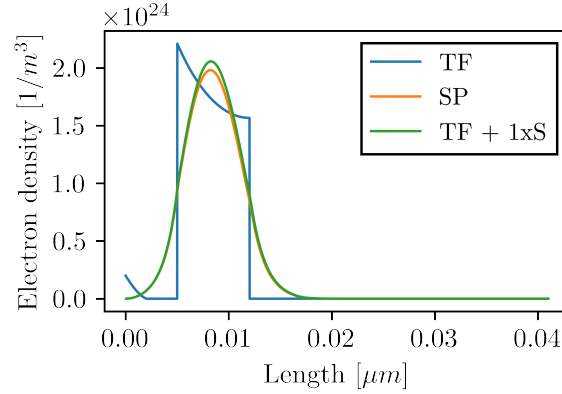
Figure 6.12: Electron density of the simple wire model, plotted as a function of distance from the wire. Semiconducting stack used here is from Fig. 6.10, simulated with a grounded aluminium wire, using $\phi_{\mathrm{w}} = 0.0325$ eV.

From the comparison in Fig. 6.12 we see that actually, the "TF + 1xS" approach is quite close to the full SP method. Thus if one has a sufficiently small system that a single diagonalization of the Hamiltonian can be calculated in a reasonable time scale, one might want to just use the "TF + 1xS" approach instead of a full SP method. However, even a single diagonalization of the Hamiltonian might itself be computationally very expensive, and thus computationally cheaper methods are still of great interest.

6.4   Calibration of λ_{vW}

Applying the SP, TF, and ETF methods to a similar model to the simple wire model in Fig. 6.12, but using the semiconducting stack from Fig. 6.2, we can compare the three methods. Such a comparison can be seen in Fig. 6.13, where we applied the ETF method for several different values of λ_{vW}.
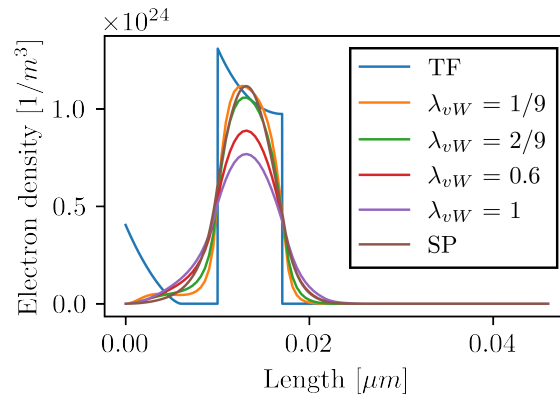

Figure 6.13: Electron density of the simple wire-region, plotted as a function of distance from the wire. Semiconducting stack used here is from Fig. 6.2, simulated with a grounded aluminium wire, using $\phi_{\mathrm{w}} = 0.058$ eV.

In Fig. 6.13 we see how the TF graph (blue) has some of the same characteristics as the SP graph (brown). For instance, the TF method generally has the same order of magnitude of the electron density as the SP method, and it also predicts electrons closely confined to the well. However, the SP method predicts a much smoother density, which underlines the weakness of the TF method. We did expect the SP method to be superior to the TF method in terms of quality and predictive power, so this is perhaps not a big surprise.

One thing that is very interesting about Fig. 6.13 is how close the ETF method is to the SP method. On the plot we clearly see that $\lambda_{\text{vW}} = 1/9$ (orange) seems a much better parameter than $\lambda_{\text{vW}} = 1$ (green), as the $\lambda_{\text{vW}} = 1/9$ graph is very similar to the SP graph. This is interesting as we from this simulation could expect to get almost as good results with the ETF method as with the SP method when using $\lambda_{\text{vW}} = 1/9$, but with a much lower computation time. We also see that in this particular simulation, the ETF method is superior to the TF method and that the ETF method also fixes the unphysical behaviour of the sharp cutoffs in the TF method.

As the barrier material in Fig. 6.13 is $In_{0.75}Ga_{0.25}As$, we have $\phi_{\text{w}} = 0.058$ eV, following the convention from section 2.4.1. With this offset between the aluminium Fermi level and the barrier conduction band minimum, we see that the electrons are quite closely confined to the wire. Going beyond this, we can look at what happens if we increase ϕ_{w}, thus filling the system with even more electrons. Using $\phi_{\text{w}} = 0.058$ eV$+$ 0.2 eV we get the band bending shown in Fig. 6.14, where the potential is calculated from the TF method.
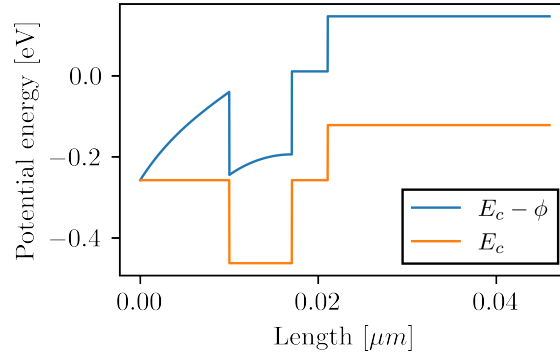

Figure 6.14: Conduction band minimum (orange), as well as conduction band minimum subtracted the electric potential (blue) of the semiconducting stack in Fig. 6.2, potential calculated from the TF method. Plotted as a function of distance from the wire. Simulated as the wire model, but here using $\phi_{\text{w}} = 0.058$ eV $+ 0.2$ eV.

We clearly see how the wire bends the bands more, thus allowing for more electrons, compared to the $\phi_{\text{w}} = 0.058$ eV case in Fig. 6.7. The corresponding electron density calculated with the TF, ETF, and SP methods are shown in Fig. 6.15.
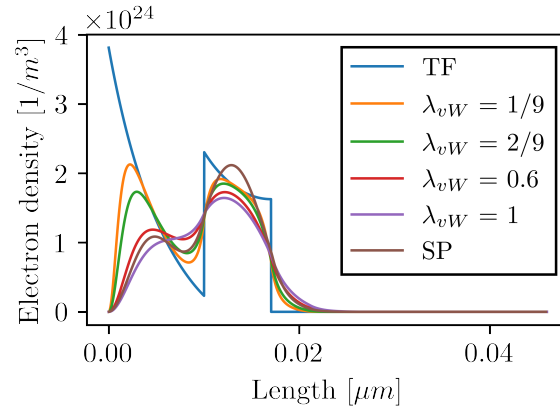
Figure 6.15: Electron density of the simple wire model as a function of distance from the wire. Semiconducting stack used here is from Fig. 6.2, simulated with a grounded aluminium wire, using $\phi_{\text{w}} = 0.058$ eV $+ 0.2$ eV.

In Fig. 6.15 we clearly see how lowering λ_{vW} approaches the steep TF density, and how raising λ_{vW} smears out the electron density. However, we also see that the $\lambda_{\text{vW}} = 1/9$ graph that seemed to fit quite well the SP curve in Fig. 6.13, now seems more off in Fig. 6.15. In Fig. 6.15 one might conclude that $\lambda_{\text{vW}} = 0.6$, or any value for λ_{vW} close to 0.6, might be a better fit. This might seem strange or perhaps even unphysical, as the parameter in the theory varies with the geometry. However, we should recall that the theory took its starting point in slowly varying densities and that λ_{vW} generally is a function of N. Thus changing ϕ_{w}, i.e. changing the conduction band minimums or applied voltage, we effectively change how many electrons are allowed in the system and affect the spatial varying of the electron density. Thus, we should judge what value of λ_{vW} fits better to each of the physical systems.

6.4.1   Introducing λ_{vW} metrics

From the above considerations, it is clear that the ETF method offers a λ_{vW} dependent approximation to the SP method. Here we will address the λ_{vW}-dependence and try to quantify the quality of each value of λ_{vW}. We will do this by introducing some metrics describing the error with respect to the SP method.

The first metric we will introduce is simply the difference in number of electrons. We will first introduce N as the number of electrons, and then define δN as the difference with respect to the SP method:

$$N = \int n(x) \, dx, \tag{6.2}$$

$$\delta N = \frac{N_{\text{vW}} - N_{\text{SP}}}{N_{\text{SP}}}. \tag{6.3}$$

Here $n(x)$ is the electron density, and x is the coordinate going through the semiconducting stack.

On top of this, we will define an error metric dealing with the difference in shape of the function. We will define δn as the difference in density squared, integrated over the semiconducting stack:

$$\delta n = \int \left(n_{\text{SP}}(x) - n_{\text{vW}}(x)\right)^2 dx. \tag{6.4}$$

On top of this we will define another metric very similar to Eq. (6.4), but with each density divided by the corresponding total number of electrons:

$$\overline{\delta n} = \int \left(\frac{n_{\text{SP}}}{N_{\text{SP}}} - \frac{n_{\text{vW}}}{N_{\text{vW}}}\right)^2 dx. \tag{6.5}$$

Along with this, we will also calculate the standard deviation. This is calculated as the standard deviation of the residuals with respect to the SP-density, i.e. it is the standard deviation of $n_{\text{SP}} - n_{\text{vW}}$.

λ	N [1/m^2]	δN	δn [1/m^5]	$\overline{\delta n}$ [1/m^5]	σ [1/m^3]
0 (TF)	8.53e+15	0.0597	1.24e+39	1.67e+07	1.64e+23
1/9	7.68e+15	-0.0460	1.28e+38	2.50e+06	5.17e+22
2/9	7.50e+15	-0.0687	4.88e+37	7.67e+05	2.99e+22
3/9	7.37e+15	-0.0841	5.76e+37	2.31e+05	3.18e+22
0.5	7.21e+15	-0.105	1.33e+38	3.63e+05	5.01e+22
0.6	7.11e+15	-0.117	1.94e+38	6.90e+05	6.12e+22
0.7	7.01e+15	-0.129	2.61e+38	1.11e+06	7.14e+22
0.8	6.91e+15	-0.141	3.32e+38	1.57e+06	8.07e+22
0.9	6.82e+15	-0.153	4.04e+38	2.05e+06	8.93e+22
1.0	6.72e+15	-0.165	4.77e+38	2.54e+06	9.71e+22
SP	8.05e+15	0	0	0	0

Table 6.3: All the metrics defined in Eq. (6.2)-(6.5) along with the standard deviation, calculated from the electron density shown in Fig. 6.13.

In Tab. 6.3 we see all the metrics calculated for the density shown in Fig. 6.13. We see for instance that $\lambda_{\text{vW}} = 1/9$ has the lowest δN, $\lambda_{\text{vW}} = 2/9$ has the lowest standard deviation, and $\lambda_{\text{vW}} = 3/9$ has the lowest $\overline{\delta n}$.

In Fig. 6.16 we plot the absolute value of δN (blue, left axis) and δn (orange, right axis) as a function of λ_{vW}. Note here that $\lambda_{\text{vW}} = 0$ corresponds to the TF method. We see that $\lambda_{\text{vW}} = 1/9$ gives the lowest error for δN and $\lambda_{\text{vW}} = 2/9$ gives the lowest error for δn.
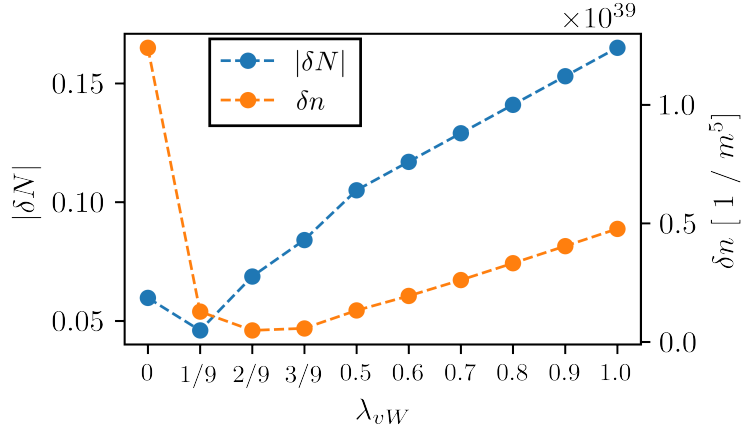
Figure 6.16: Plot of the absolute value of δN (blue, left axis) and δn (orange, right axis) as a function of λ_{vW}. Data from Tab. 6.3.

Instead of considering the stack in Fig. 6.2, we will now take a look at the stack of the dot-probe, as can be seen in Fig. 6.10. As one can see, the materials are different here, and the barrier is only 5 nm thick, in contrast to the 10 nm barrier of the stack in Fig. 6.2. Making a similar simulation to the one in Fig. 6.15, but using the dot-probe stack from Fig. 6.10, we get the results as can be seen in Fig. 6.17.
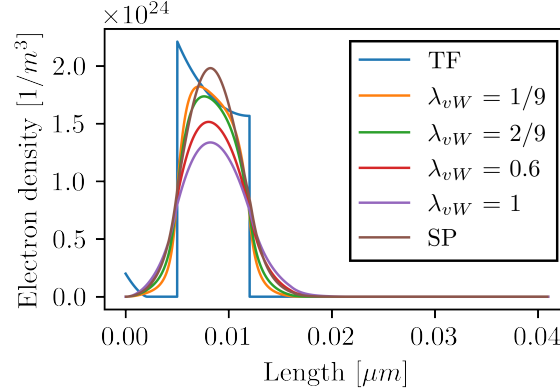

Figure 6.17: Electron density of the simple wire model as a function of distance from the wire. Semiconducting stack used here is from Fig. 6.10, simulated with a grounded aluminium wire, using $\phi_{w} = 0.0325$ eV.

Here we see that $\lambda_{vW} = 1/9$ again seems like a good fit, and that it seems to fit better than $\lambda_{vW} = 2/9$, or any other of the curves.

6.5   ETF based simulations

In this section, we will apply the ETF method to the dot-probe device shown in Fig. 6.10. We will use $\lambda_{vW} = 1/9$ here, as this is found as an appropriate value from the

results in section 6.4. Applying the ETF method to the dot-probe device, while keeping the gate voltages the same as in Fig. 6.11, yields the electron density as can be seen in Fig. 6.18 (a).

(a) Electron density ($\lambda_{vW} = 1/9$) [1/m^3].   (b) Electron density ($\lambda_{vW} = 0.5$) [1/m^3].
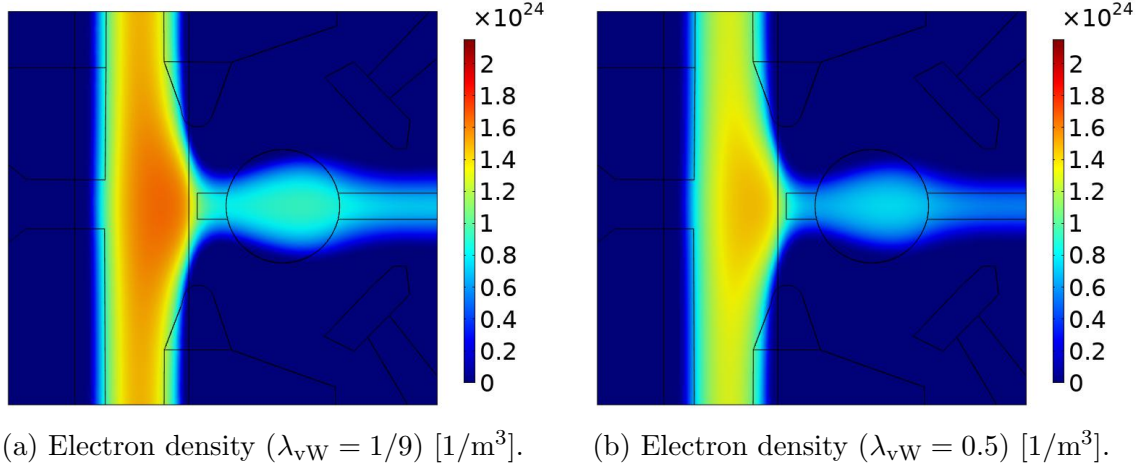
Figure 6.18: Plots of the electron density with $\lambda_{vW} = 1/9$ (a) and with $\lambda_{vW} = 0.5$ (b) for the dot-probe device. Plots are made in the middle of the InAs well.

We added the $\lambda_{vW} = 0.5$ case to Fig. 6.18 (b) for reference. Here we see how the electron density in 6.18 (a) is very similar to the electron density from the TF method in Fig. 6.11. As is expected, 6.18 (b) shows how increasing λ_{vW} smears out the electron density, and thus it is also expected that the $\lambda_{vW} = 1/9$ case in Fig. 6.18 (a) is somewhat similar to the TF method in Fig. 6.11. If we take a look at the error defined in Eq. (6.1) for the two cases in Fig. 6.18, we get the results shown in Fig. 6.19
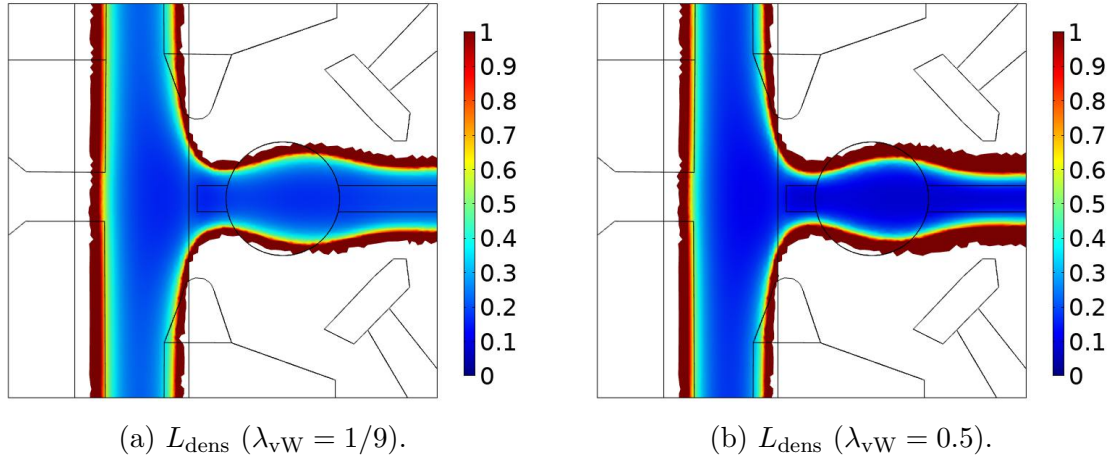

(a) L_{dens} ($\lambda_{vW} = 1/9$).   (b) L_{dens} ($\lambda_{vW} = 0.5$).

Figure 6.19: Plots of L_{dens} [unitless] with $\lambda_{vW} = 1/9$ (a) and L_{dens} with $\lambda_{vW} = 0.5$ (b) as described in Eq. (6.1) for the dot-probe device. Color scale of L_{dens} cut at a maximum of 1. Both plots are made in the middle of the InAs well.

Here we again added the $\lambda_{vW} = 0.5$ case in Fig. 6.19 (b) for reference. Comparing

L_{dens} (from the definition in Eq. (6.1)) for the $\lambda_{\mathrm{vW}} = 1/9$ case in figure 6.19 (a) to the TF method in Fig. 6.11 (b), we do see a subtle but clear difference. In Fig. 6.19 we also see how increasing λ_{vW} smoothens out the electron density, making it more slowly varying. One should note that L_{dens} in Fig. 6.19 is not plotted in the physically forbidden regions where the potential is below the conduction band minimum.

6.5.1   Quantum dot test case

To further test the ETF method, we have made a toy model resembling many aspects of the dot-probe device. This model has the same stack as the dot-probe device, i.e. the stack in Fig. 6.10, and all the same material properties. The geometry of this model can be seen in Fig. 6.20.
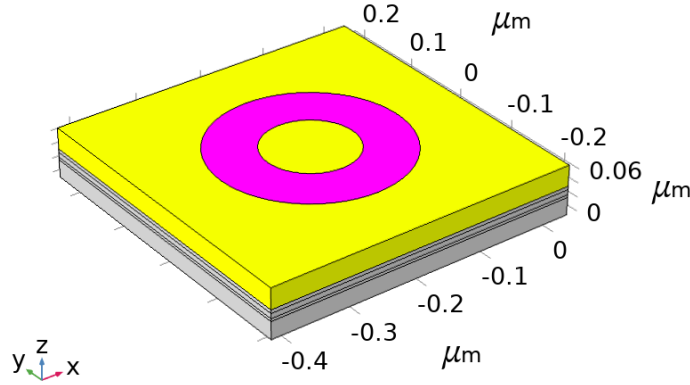

Figure 6.20: Simplified dot-device with inner gate of 65 nm radius.

On top of the semiconducting stack, we placed two gates (yellow in the figure) separated with a ring of oxide (HfO$_2$, purple in the figure). This means that in the toy model, there is no oxide separating the gates and the semiconductor; there is only oxide between the gates. The inner gate is the circle in Fig. 6.20, the oxide is the ring around it, and the outer gate is the exterior region surrounding the ring. The gates are modelled as aluminium, (i.e. we use $\phi_{\mathrm{w}} = 0.0325$ eV), and the inner gate has a radius of 65 nm. Here we should note that aluminium has been in the form of a grounded wire so far in this thesis. However, we will consider it as a gate for this toy model, i.e. we will apply non-zero voltages to the gates.

We then swept over the voltage of the inner and outer gates and found that when both gates were around -0.35 V, we had perfect depletion in the stack. Fixing then the outer gate at -0.35 V, we swept the voltage of the inner gate to fill the area under the inner gate with electrons. To compare the TF and ETF methods more quantitatively, we integrated the charge in the semiconducting stack for each inner gate voltage and thus got how many electrons each voltage gave rise to for both the TF and ETF methods. We did this for three different sizes of the inner gate, a radius of 65 nm (as can be seen in Fig. 6.20), as well as a radius of 30 nm and 100 nm. The results can be seen in Fig. 6.21.
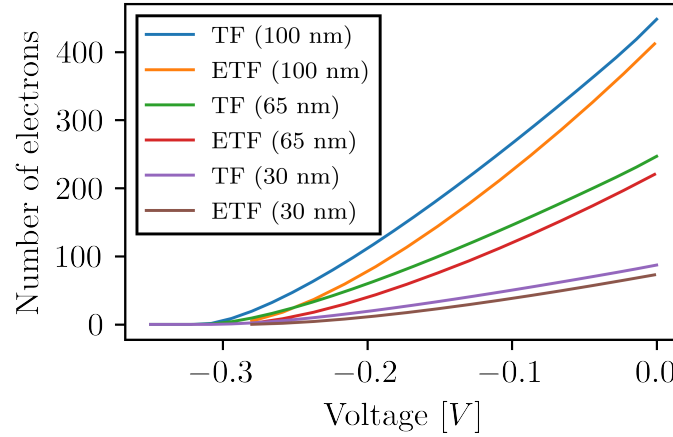
Figure 6.21: Number of electrons in the semiconducting stack as a function of inner gate voltage. Shown for the TF and ETF methods, for three different inner gate radii.

In Fig. 6.21 we clearly see how the TF method predicts more electrons than the ETF method. The same can be seen in Tab. 6.3, where it is clear that the TF method predicts the largest number of electrons when compared to the ETF and SP methods. Running the 65 nm test case with a larger voltage range and also applying the SP method, we get the results, as can be seen in Fig. 6.22
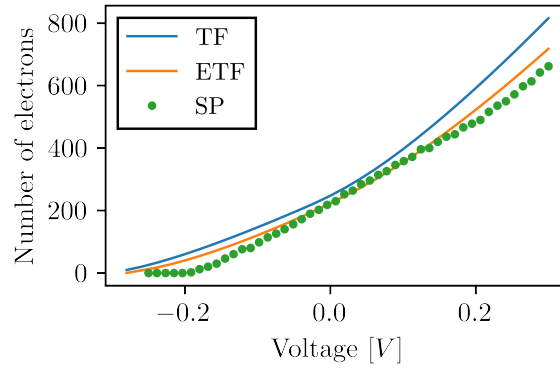

Figure 6.22: Number of electrons in the semiconducting stack as a function of inner gate voltage. Shown for the TF, ETF, and SP methods, for an inner gate radius of 65 nm.

In Fig. 6.22 we see how the SP method generally predicts depletion at a larger voltage than the TF and ETF methods. We also note how the SP method generally seems to better match with the ETF method. However, one should be very cautious with the SP results in Fig. 6.22. It was computationally very time-consuming to run the SP simulation, and thus we used a somewhat rough grid. This grid had discretization of 2.5 nm in x- and y- direction and discretization of 0.9 nm in the z-direction. This is rougher than our TF and ETF simulations, and thus we will not use the SP results in Fig. 6.22 for further calculations.

Taking the TF and ETF results in Fig. 6.22 and calculating the capacitance as dq/dV, where q is the charge in the stack, we get the results shown in Fig. 6.23.
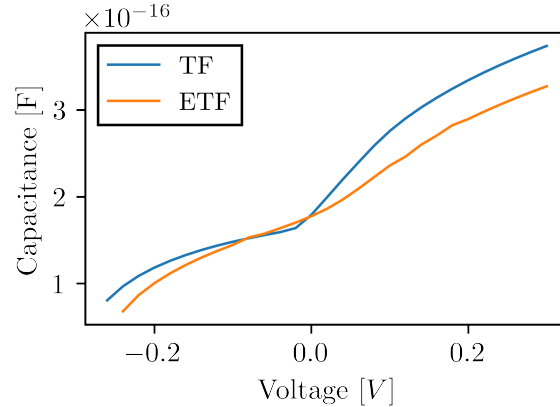

Figure 6.23: Capacitance as a function of gate voltage, shown for the TF and ETF methods with an inner gate radius of 65 nm.

In Fig. 6.23 we see how the TF predicted capacitance has a kink around 0 V. This happens as the TF method has regions that are forbidden at certain voltages, but as you increase the voltage, these regions will at some point stop being forbidden. Thus at a gate voltage of 0 V, other regions are available for the electrons than at a gate voltage of say -0.15 V. For the ETF method, this is not the case. Here there are no forbidden regions in the semiconducting stack, and therefore the approach predicts "tails" of electron density out in the regions forbidden for the TF method. Therefore these regions slowly fill up as you increase the gate voltage, in contrast to the sharp transition of the TF method. This difference is very clear in Fig. 6.23, and one can argue that this TF behaviour is unphysical. It would be very interesting for future work to do the same capacitance simulation using the SP method and compare this with the TF and ETF results. However, it requires big computational power to run the SP method on a reasonably fine mesh.

6.6   Convergence, mesh, and solver configurations

Throughout this thesis, we have made several simulations using different setups, approaches, and even different software. Below we seek to give a general overview of the numerical details and give some insight into the general setup and line of thought.

Naturally, all the models have been applied a mesh or grid. This meshing turns out to be important for both the quality of the results and the convergence of the system. In the 1D systems (3D with 2 directions that are translational invariant, thus effectively 1D) studied with Python code, we have made a simple grid going through the semiconducting stack. The spacing between each grid point in this lattice varies with system and approach. For instance, we can make the grid of the TF method much finer than for the SP method simply because of the speed of the TF method. Generally

speaking, we have in all these cases kept a distance of each grid point of less than ~ 0.1 nm. This is not entirely arbitrarily chosen. Looking at the Fermi wavelength of aluminium for instance, $\lambda_F = 2\pi\hbar/\sqrt{2mE_F} \approx 0.36$ nm [10], we wish to be below that value. We do not solve the electrostatics inside the aluminium, and we would expect the wavelength to be larger in the semiconductor [10], so a grid spacing of maximum ~ 0.1 nm should offer sufficient resolution. For the TF and vW cases, we often use a spacing of ~ 0.01 nm, simply because speed is not an issue here.

For the full-scale device simulations, the meshing is more challenging. These simulations are made in COMSOL Multiphysics, which has several advanced meshing options. This allows one to mesh the regions of great interest finely and other regions coarsely. For instance, we have meshed the well and the area under the wire finely. In Fig. 6.24 we see the meshing of the dot-probe device. One sees how the upper parts are coarsely meshed, but the well is very finely meshed (the well looks all black for all the meshing points).
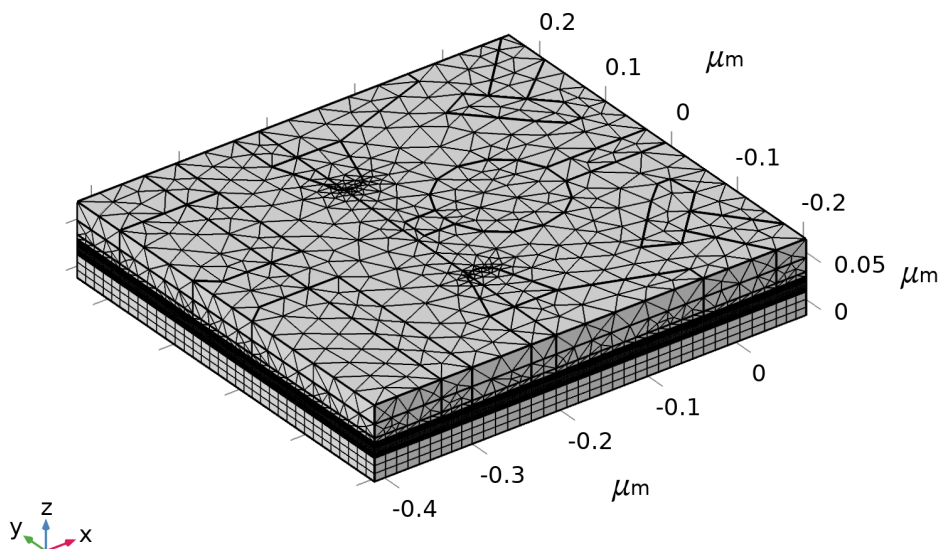

Figure 6.24: Meshing of the dot-probe device.

The particular mesh in Fig. 6.24 is made in COMSOL Multiphysics, using built-in options to customize the mesh. In the upper parts of the geometry, i.e. in the oxide and gate regions, the so-called free tetrahedral meshing has been used. This is a very agile meshing that automatically handles curves, such as the ones we have on the gates. In the semiconducting region, we have used a so-called swept mesh that allows us to specify the distribution of mesh elements. Thus we have specified a very fine meshing around the well and, for instance, a very coarse mesh in the sub-barrier. This is a clever way of meshing, as the computational task of simulating the entire device with meshing

as fine as in the well would be very time-consuming, and most of the device is not an area of great interest anyway.

Generally speaking, we have had a very robust and fast convergence when applying the TF method. However, the ETF method has been more challenging to implement. We have applied the ETF method through the weak form described in section 5.2.1, and applied it in COMSOL Multiphysics using the Weak Form PDE module. Poisson's equation was then introduced using the Electrostatics module, thus closing the loop of the self-consistent problem. We tried several different solver configurations and found a relatively robust convergence using a direct solver, but this configuration was rather slow. We could speed up the problem using an iterative solver, with the coupling between Poisson's equation and the ETF functional through a segregated solver. For the TF and ETF simulations, we used a relative tolerance of 10^{-5} (as defined in Eq. (5.8)), and for the iterative solver, we primarily used a so-called generalized minimal residual method (GMRES). Along with this, we used a geometric multi-grid that speeds up the simulations even further. We will not discuss the solver configuration in greater detail but instead refer to the COMSOL Multiphysics documentation for further details about how the different solvers work. This documentation can be found in Ref. [32].

Along the way, we did as many checks of the results as possible. For instance, we made simple 1D models where we applied TF and SP methods and checked these results by calculating them independently by hand. For instance, the SP method predicts simple harmonics for the particle in a box textbook example, which is a great way to check that the code is working as intended. We also reproduced results from other papers, for instance, the TF results in Ref. [4]. In this way, we could check the results along the way, to see that we were on the right track.

6.7 Applications of simulation methods

In the sections above, we have investigated and analysed 3 different numerical approaches (TF, ETF, and SP), and applied these to different 2DEG-based devices in order to simulate electric potential and electron density. In this section, we have applied some of the numerical methods to different devices, and not only with the sole purpose of modelling electric potential and electron density. On top of this, we have included some of the physical effects that we have purposely excluded until now. These effects are for instance surface charge between the oxide and the barrier, along with holes in the semiconductors.

This section will not discuss the details in great depth, but rather keep this as an overview to show some of the opportunities available with the numerical approaches.

6.7.1 Nanowire geometry

The numerics above can also be applied to another very common geometry, the nanowire geometry. In the following, we will consider a hexagonal nanowire consisting of InAs, shown in Fig. 6.25
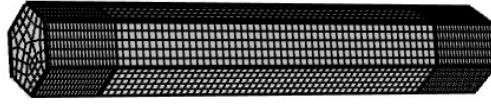
Figure 6.25: Model of the the hexagonal nanowire, shown here with the applied mesh.

We will split the nanowire into three different parts. Two end parts of length 100 nm, shown with a relatively fine meshing in Fig. 6.25, and one middle part of length 500 nm, shown with a relatively coarse meshing in Fig. 6.25. On the middle part, we will apply the electric potential of a grounded aluminium gate on two adjacent sides out of the 6. We will also apply an electric potential through aluminium gates at the two ends of the wire. Note here that we again treat the aluminium parts as gates and not as grounded wires. All sides that are not applied aluminium gates are applied with a constant surface charge of $1.3 \cdot 10^{16}$ e/m^2, where e is the electron charge. Along with this, we will assign an offset between InAs and aluminium of $\phi_\text{w} = 0.35$ eV. These numbers can be found in Ref. [43], where the geometry is also further discussed and justified. In Fig. 6.26 we see the electric potential (a) and electron density (b) through the middle of the nanowire, calculated with the ETF method. We used the Dirichlet boundary condition of $\psi = 0$ for all edges in the system, as can also be seen in Fig. 6.26 (b), where the electron density is zero at the ends of the wire.

(a) Electric potential along the wire.
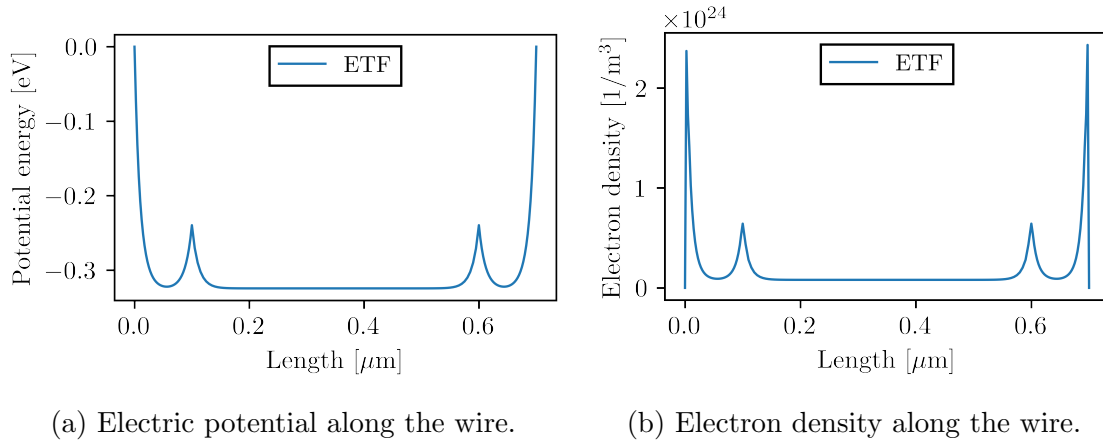

(b) Electron density along the wire.

Figure 6.26: Results are calculated along the middle of the wire with the ETF method using $\lambda_\text{vW} = 1/9$.

We have also simulated a cross-section in the middle of the wire. This is made as a 2D simulation, using a very fine meshing on the cross-section, as shown in Fig. 6.27.
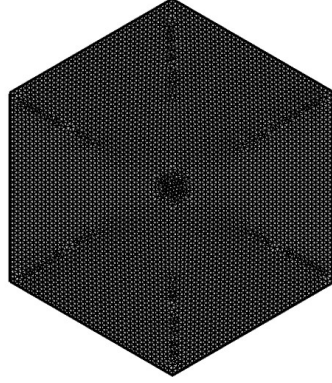
Figure 6.27: Cross-section of the the hexagonal nanowire, shown here with the applied mesh.

Again we used a Dirichlet boundary condition of $\psi = 0$ for the ETF method. The electron density of this simulation can be seen in Fig. 6.28 for both the TF (a), and ETF (b) methods.
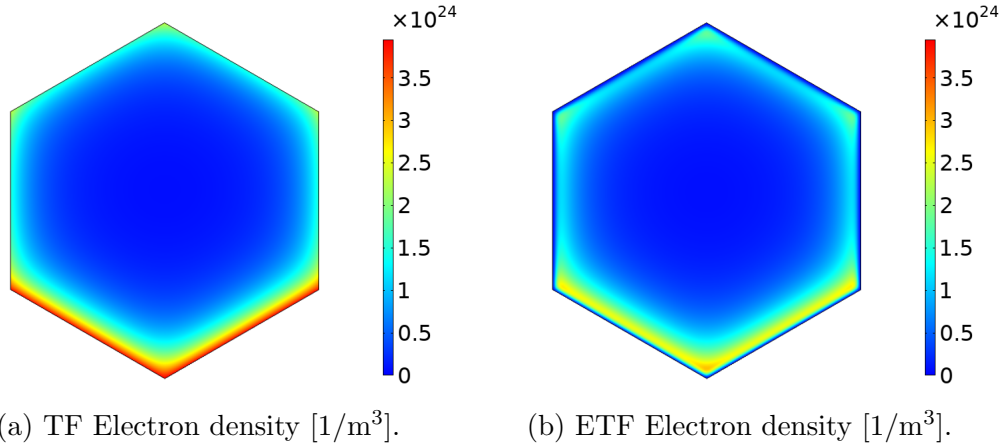

(a) TF Electron density $[1/\mathrm{m}^3]$.

(b) ETF Electron density $[1/\mathrm{m}^3]$.

Figure 6.28: Electron density of the cross-section of the nanowire. Shown for the TF method (a) and ETF method with $\lambda_{\mathrm{vW}} = 1/9$ (b).

Again we clearly see how $\psi = 0$ is applied to the edges for the ETF method in Fig. 6.28 (b). In Fig. 6.28 we applied the aluminium gates to the two lower sides of the hexagonal cross-section and applied the surface charge to the four other sides. One clearly sees how the aluminium attracts a larger amount of electrons on the two aluminium sides than on the four surface charge sides. From these results, we again see how the TF method predicts a larger electron density compared to the ETF method.

6.7.2   Surface density and holes

Except for the nanowire simulations above (that assumed a constant surface charge), we have not run any simulations including surface charge between the semiconducting stack and the oxide. As is for instance discussed in Ref. [44], we do expect a non-negligible positive surface charge density. We will model this surface charge density as

a 2D TF density, as is derived in Eq. (4.30). This means that we will generally model the surface charge as

$$\sigma_{\mathrm{s}} = eD_{\mathrm{is}}(-eV - \phi_{\mathrm{w}} + \phi_{\mathrm{CNL}}), \qquad (-eV - \phi_{\mathrm{w}} + \phi_{\mathrm{CNL}}) > 0. \tag{6.6}$$

Here e is the electron charge, V is the electric potential, ϕ_{w} is the offset between the conduction band minimum of the barrier and the Fermi level of the aluminium, D_{is} is the density of interface states, ϕ_{CNL} is the charge neutrality level. For the 2DEG simulations, we will assume the values $D_{\mathrm{is}} = 4 \cdot 10^{16}$ $1/\mathrm{m}^2$ and $\phi_{\mathrm{CNL}} = 300$ meV. These are somewhat arbitrarily chosen, but are similar to the values in the literature, see for instance Ref. [43, 44].

In the simulations below, we will also consider holes. We will treat the holes as positive electrons, i.e. describe them using the 3D TF density, as described in Eq. (4.31). We will consider both heavy holes and light holes and model the hole density as a sum of the two. Thus we will model the charge density caused by holes as

$$\sigma_{\mathrm{h}} = \frac{e}{3\pi^2}\left(\left(\frac{2m_{\mathrm{lh}}}{\hbar^2}\right)^{3/2} + \left(\frac{2m_{\mathrm{hh}}}{\hbar^2}\right)^{3/2}\right)(-eV - \phi_{\mathrm{w}} + E_v)^{3/2}, \qquad (-eV - \phi_{\mathrm{w}} + E_v) > 0. \tag{6.7}$$

Here m_{lh} is light hole mass, m_{hh} is the heavy hole mass, and E_v is the valence band maximum. When taking the holes into account, Eq. (6.7) will be applied to the entire semiconducting stack.

6.7.3 Spin-orbit field

We have also considered the effect of spin-orbit field, as this is essential for hybrid devices as well. Following Ref. [45], we can calculate the Rashba coupling from the average electric field of the semiconductor. Please note that Ref. [45] takes its point of reference in an eight-band $\mathbf{k} \cdot \mathbf{p}$ model, which we have not discussed in this thesis. This eight-band Kane model offers a more detailed description of the band structure of the semiconductors, but is thus a more complicated model. We will not go into greater detail about the Kane model but instead refer the interested reader to Ref. [17]. The Rashba coupling can thus be calculated as [45]

$$\alpha = \frac{eP^2}{3}\left(\frac{1}{E_0^2} - \frac{1}{(E_0 + \Delta_0)^2}\right)\varepsilon. \tag{6.8}$$

Where ε is the averaged electric field. We used the same material properties for InAs as Ref. [45], and calculated the values for the other materials through linear interpolation. These properties are shown in Tab. 6.4.

Property	InAs	In$_{0.82}$Al$_{0.18}$As	In$_{0.81}$Al$_{0.19}$As	In$_{0.75}$Ga$_{0.25}$As
P [eV nm]	0.9197	0.9156	-	0.9521
E_0 [eV]	0.418	0.906	-	0.693
Δ_0 [eV]	0.380	0.366	-	0.370
m_{lh}/m_e	0.026	0.06	0.06	0.04
m_{hh}/m_e	0.41	0.42	0.42	0.42

Table 6.4: Relevant physical properties used for the Rashba coupling simulations. Material properties collected from several references, see Ref. [39, 45, 46, 47].

In Ref. [45] they consider a nanowire geometry, and thus the averaging over electric field is naturally different from the averaging in 2DEG devices. We will make a toy model with characteristics of the usual 2DEG devices and plot the Rashba coupling as a function of the magnitude of the electric field. The model we will consider is a simplified gate geometry, using the semiconducting stack of Fig. 6.2, but applying here a 50 nm sub barrier i.e. 50 nm of In$_{0.82}$Al$_{0.18}$As. This is schematically shown in Fig. 6.29, using the same colour code as in the stack in Fig. 6.2.
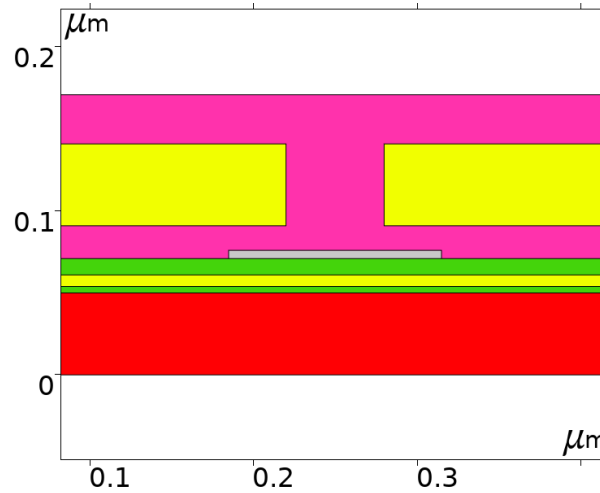

Figure 6.29: Schematic of the simplified geometry. Colour code following Fig. 6.2.

In Fig. 6.29 we see that the simplified model is a cross-sectional cut going through the aluminium wire (grey in the figure). We see the semiconducting stack (red, green, yellow and green), the HfO$_2$ oxide (purple), as well as two golden gates (yellow). We applied a gate voltage of -4 V and grounded the aluminium wire. We also applied the charge of holes to the semiconducting stack and a surface charge between the semiconducting stack and the oxide, as described in section 6.7.2. This gave rise to the electron density, as shown in Fig. 6.30.

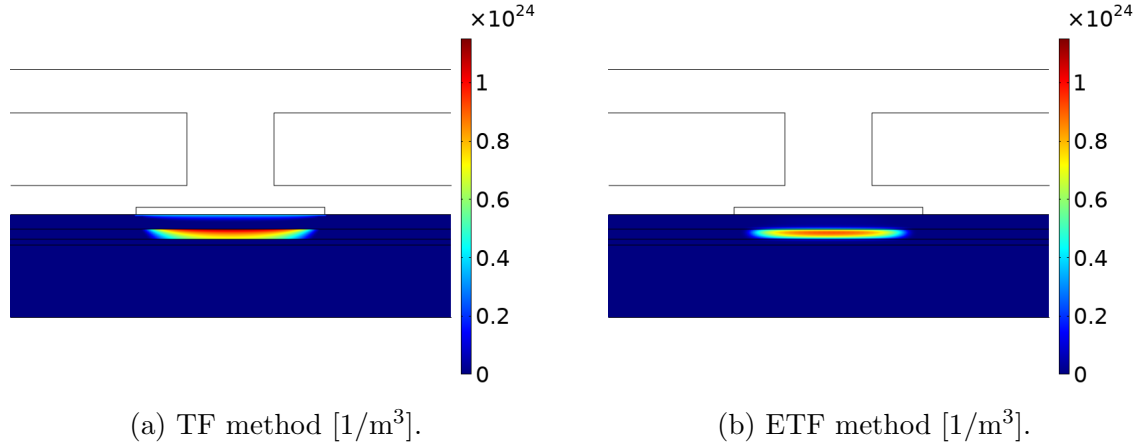(a) TF method [1/m^3].   (b) ETF method [1/m^3].

Figure 6.30: Electron density of the simplified model. Shown for TF method (a), and for the ETF method with $\lambda_{vW} = 1/9$ (b).

In Fig. 6.30 we clearly see how the gates induce perfect depletion everywhere except for the area underneath the wire. Note how the ETF method (6.30 (b)) predicts a smooth transition of electron density going from the filled well to the depleted barriers, whereas the TF method (6.30 (a)) predicts a sharp behaviour. Plotting now the Rashba coupling using the magnitude of the electric field, as shown in Eq. (6.8) and ETF method for the electron density, we get the plot shown in Fig. 6.31.
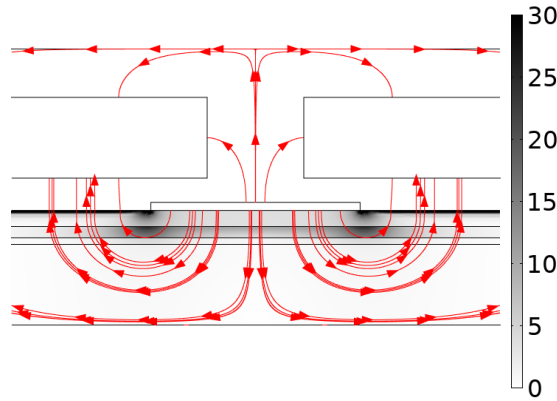

Figure 6.31: Rashba coupling plotted on the colour scale [meV nm] as described in Eq. (6.8), and electric field lines with direction plotted as red arrows. Calculated with the ETF method using $\lambda_{vW} = 1/9$.

In Fig. 6.31 we have the magnitude of the Rashba coupling shown on the colour scale. We have also included the electric field lines and their direction, as shown on the red arrows. We see naturally how the electric field vanishes inside the gates and the wire and how the sub-barrier has a Rashba coupling of 0.

Chapter 7

Conclusion and outlook

This thesis has investigated different methods for electrostatic simulations of nanoelectronic hybrid devices. We have particularly focused on 2DEG Majorana devices.

We started by considering a simplified model resembling the MOS part of a hybrid device, i.e. gate, oxide, barrier, and well. Applying Gauss' law along with appropriate boundary conditions, we analytically described the evolution of the electric field and electric potential through the device. We did this by taking several different charge densities into account, including a constant oxide charge, an oxide-barrier surface charge, and a charge density in the well. In the end, we solved for the charge density in the well, as this is the area of interest for the experimental physicists working on the devices. With the analytically calculated electron density in the well, we could compare to real experimental data of the electron density measured from a Hall bar experiment. Here, the intention was to calibrate the model to the experimental data and gain confidence in the model's predictive power. However, we found that the model could not fit the experimental data. This is shown in Fig. 3.5.

Several reasons could cause the fact that the model could not recreate the experimental data. One likely reason is that the basic assumption in the form of g_{w} (the ansatz we made) is not sufficiently good and that the simplified geometry is too crude to approximate the real device. It would definitely be interesting to follow this line of thought and try to get an analytical expression for the charge accumulation in the well as a function of gate voltage. This could be very useful since one could then calibrate the model, which is an essential part of the modelling process. However, this might require a very advanced analytical model.

To numerically describe the electrostatics, one needs to couple Poisson's equation with some equation describing the electron density as a function of electric potential. In this way, one can solve Poisson's equation for the electric potential and use this in the other equation to solve for the electron density, which again can be used in Poisson's equation to solve for the next iteration of electric potential. In this way, one can describe the electrostatics through a self-consistent set of equations. However, several different approaches can be taken to obtain the equation that couples with Poisson's equation. This thesis has investigated three different approaches to close the self-consistent loop.

The Schrödinger-Poisson (SP) method is a widely used approach that has the advantage of producing very good results compared to the other approaches discussed in this thesis. However, the SP method is very computationally costly because it solves Schrödinger's equation for the wave function of the electrons. This means that diagonalization of the Hamiltonian is needed, which is a computationally very costly operation. This is further described in section 5.1.2. With the wave function from Schrödinger's

equation, one can easily calculate the electron density and thus close the self-consistent loop, as is described in Fig. 5.2.

The Thomas-Fermi (TF) method is another widely used method that has the advantage of being computationally very fast. The TF method is an orbital-free (OF) density functional theory, meaning that it completely bypasses the wave functions of the system and instead locally treats the electron density as the basic variable. This is clever, as it thus removes the need to diagonalize the Hamiltonian in Schrödinger's equation. One can derive the TF electron density by assuming a slowly varying electron density, as is done in section 4.3. One can use this to close the self-consistent loop, as is illustrated in Fig. 4.1.

The extended Thomas-Fermi (ETF) method is also an OF density functional theory and is constructed very similarly to the TF method. Actually, it consists of the same energy functional but has the von Weizsäcker (vW) term added. The vW term takes spatial modulation of the electron density into account by including the gradient of the electron density. Thus it tends to smear out the electron density more than the TF method. This is, in many cases, an improvement to the TF method, as the TF method seems to have a somewhat unphysical behaviour of very rapidly changing electron density. The ETF method is derived in sections 4.4 and 5.2. Here we use a minimisation of the weak formulation to close the self-consistent loop. This is sketched in Fig. 5.3.

With these three numerical methods, we were able to simulate real devices. In section 6.1 and 6.2 we apply the TF method to several real geometries and investigate the weaknesses of the TF method. Here we see, that the TF is generally good as a first approximation but fails on several parameters when taking a closer look. In section 6.4 we calibrate the vW coefficient and compare the three methods (SP, TF, and ETF). This is, for instance, done on a real semiconducting stack in Fig. 6.13. Here, we conclude that generally $\lambda_{vW} = 1/9$ is a good value for our devices and compare the three methods, using $\lambda_{vW} = 1/9$ for the ETF case. Here we also see how the TF method fails with respect to the SP method and how the ETF method is somewhere in between the TF and SP methods. We do further tests of the TF and ETF methods in section 6.5.1. Here we see a remarkable difference in capacitance when comparing the two approaches, which can be seen in Fig. 6.23.

To summarise, we have derived three different methods for simulating electrostatics of nanoelectronic devices. We found that the TF method is very fast but unphysical in some situations. In contrast to this, we have the SP method that is very slow but have excellent results. On top of this, we have constructed an extended Thomas-Fermi (ETF) method. This is another OF approach, constructed similar to the TF method, but added the von Weizsäcker term that takes spatial modulation of electron density into account. It gives results of quality that seem to lie somewhere between the TF and SP methods. The speed of the ETF method also lies somewhere between the TF and SP methods and can thus be considered a sort of midway between the two.

It would be interesting for future work to investigate the difference between the three methods further. It would be very useful to see more detailed comparisons and

determine whether the ETF method has flaws that we have not seen yet. If the ETF method really does perform well in many cases, as we have seen that it does in a few cases, it might be an approach of great interest for future nanoelectronic simulations.

Appendix A

Electrostatic calculation of the wire-model

Here we show the electrostatic calculations applied to Fig. 3.2. The system is again shown here for simplicity (just a copy of Fig. 3.2).
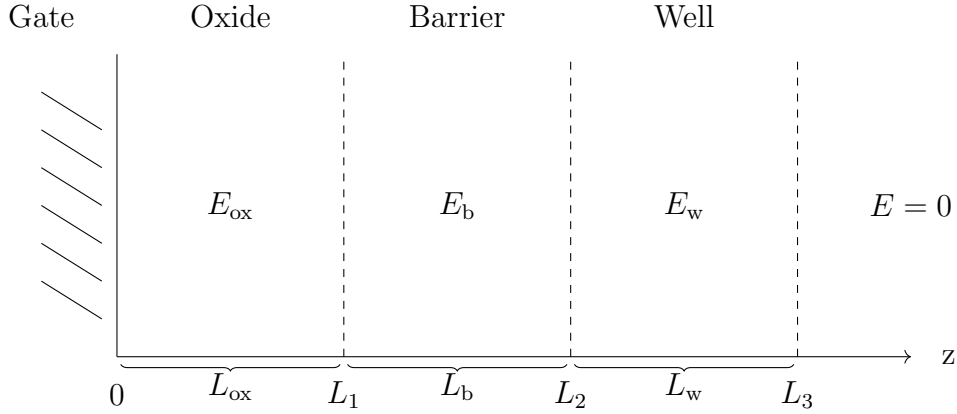

Figure A.1: Sketch of the simple 1D model.

Below we will do the calculation for the case of σ_w confined to the far back of the well, but the calculation can easily be done for any other charge configuration, for instance, the ones shown in Fig. 3.3. As is described in section 3, the only difference will turn out to be the weight α. Recall that we apply a Dirichlet boundary condition of $V = V_\mathrm{G}$ at the gate-oxide interface, along with a Neumann condition of $E = 0$ at the back end of the well. We will use the notation from the figure above and start by applying Gauss' law to the barrier-well interface and the back end of the well, where the charge is confined. We get that

$$\epsilon_\mathrm{b} E_\mathrm{b} = \epsilon_\mathrm{w} E_\mathrm{w}, \tag{A.1}$$

and

$$\epsilon_\mathrm{w} E_\mathrm{w} = -\sigma_\mathrm{w}. \tag{A.2}$$

Here ϵ is the permittivity, and E is the electric field. Along with this, we will use the assumptions of σ_w and σ_s

$$\sigma_\mathrm{w} = -e g_\mathrm{w}(e V_\mathrm{w} + \phi_\mathrm{w}), \tag{A.3}$$

and

$$\sigma_\mathrm{s} = e g_\mathrm{DIT}(-e V_\mathrm{s} + \phi_\mathrm{s}). \tag{A.4}$$

Here e is the electron charge, g_w and g_DIT are constants of the surface charge, V refers to the electric potential, ϕ_w and ϕ_s are the offsets of the conduction band minimums. We used that the electrostatic energy is $-eV$.

Along with this, we will apply Gauss' law to the barrier-oxide interface,

$$\epsilon_b E_b - \epsilon_{ox} E_{ox} = \rho L_{ox} + \sigma_s, \tag{A.5}$$

where ρ is the charge density in the oxide. We must also conclude that the electric potential at the oxide-barrier interface is given by

$$V_s = V_G - E_{ox} L_{ox}. \tag{A.6}$$

Following the same reasoning, the electric potential in the well must be given by

$$V_w = V_G - E_{ox} L_{ox} - E_b L_b - E_w L_w. \tag{A.7}$$

Using now Eq. (A.1) and (A.2) we get that

$$E_b = -\frac{\sigma_w}{\epsilon_b}, \quad \text{and} \ \ E_w = -\frac{\sigma_w}{\epsilon_w}. \tag{A.8}$$

Using Eq. (A.8) along with Eq. (A.1), we get that

$$E_{ox} = \frac{\epsilon_b E_b - \rho L_{ox} - \sigma_s}{\epsilon_{ox}} = \frac{-\sigma_w - \rho L_{ox} - \sigma_s}{\epsilon_{ox}}. \tag{A.9}$$

From Eq. (A.3) we can write the electric potential in the well as

$$V_w = -\frac{\sigma_w}{g_w e^2} + \frac{\phi_w}{e}. \tag{A.10}$$

Setting now Eq. (A.7) equal to Eq. (A.10), we get that

$$-\frac{\sigma_w}{g_w e^2} + \frac{\phi_w}{e} = V_G - E_{ox} L_{ox} - E_b L_b - E_w L_w. \tag{A.11}$$

Inserting now Eq. (A.8) and (A.9), we can rewrite Eq. (A.11) as

$$-\frac{\sigma_w}{g_w e^2} + \frac{\phi_w}{e} = V_G + \frac{L_{ox}}{\epsilon_{ox}}\left(\sigma_w + \sigma_s + \rho L_{ox}\right) + \frac{L_b}{\epsilon_b}\sigma_w + \frac{L_w}{\epsilon_w}\sigma_w. \tag{A.12}$$

We will now make the following definitions:

$$C_{ox} \equiv \frac{\epsilon_{ox}}{L_{ox}}, \ \ C_b \equiv \frac{\epsilon_b}{L_b}, \ \ C_w \equiv \frac{\epsilon_w}{L_w}, \ \ \text{and} \ \ C_{st} \equiv \left(C_{ox}^{-1} + C_b^{-1} + C_w^{-1}\right)^{-1}. \tag{A.13}$$

This allows us to write Eq. (A.12) as

$$\frac{\phi_w}{e} - V_G - \frac{\sigma_s + \rho L_{ox}}{C_{ox}} = \sigma_w\left(\frac{1}{g_w e^2} + C_{ox}^{-1} + C_b^{-1} + C_w^{-1}\right). \tag{A.14}$$

Isolating σ_w in Eq. (A.14) gives us

$$\sigma_w = \left(\frac{1}{g_w e^2} + C_{st}^{-1}\right)^{-1}\left(\frac{\phi_w}{e} - V_G - \frac{\sigma_s + \rho L_{ox}}{C_{ox}}\right). \tag{A.15}$$

Writing out C_{st}, and introducing the weight of the well, α, i.e. now considering any charge distribution in the well, we get that

$$\sigma_{\text{w}} = \left(\frac{1}{g_{\text{w}}e^2} + \alpha\frac{L_{\text{w}}}{\epsilon_{\text{w}}} + \frac{L_{\text{ox}}}{\epsilon_{\text{ox}}} + \frac{L_{\text{b}}}{\epsilon_{\text{b}}} \right)^{-1} \left(\frac{\phi_{\text{w}}}{e} - V_{\text{G}} - \frac{\sigma_{\text{s}} + \rho L_{\text{ox}}}{C_{\text{ox}}} \right). \tag{A.16}$$

Naturally, one can calculate the electron density as $n_{\text{w}} = \sigma_{\text{w}}/(-e)$

$$n_{\text{w}} = \frac{1}{e} \left(\frac{1}{g_{\text{w}}e^2} + \alpha\frac{L_{\text{w}}}{\epsilon_{\text{w}}} + \frac{L_{\text{ox}}}{\epsilon_{\text{ox}}} + \frac{L_{\text{b}}}{\epsilon_{\text{b}}} \right)^{-1} \left(-\frac{\phi_{\text{w}}}{e} + V_{\text{G}} + \frac{\sigma_{\text{s}} + \rho L_{\text{ox}}}{C_{\text{ox}}} \right). \tag{A.17}$$

Bibliography

[1] Chetan Nayak, Steven H. Simon, Ady Stern, Michael Freedman, and Sankar Das Sarma. Non-Abelian anyons and topological quantum computation. *Reviews of Modern Physics*, 80(3):1083–1159, 2008.

[2] Thomas Ihn. *Semiconductor Nanostructures - Quantum States and Electronic Transport*. Oxford University Press, 2010.

[3] Ivan H. Deutsch. Harnessing the Power of the Second Quantum Revolution. *Physical Review Applied*, 10(1):1, 2020.

[4] August E. G. Mikkelsen, Panagiotis Kotetes, Peter Krogstrup, and Karsten Flensberg. Hybridization at superconductor-semiconductor interfaces. *Phys. Rev. X*, 8, Aug 2018.

[5] Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, Oct 1997.

[6] Andrew Steane. Quantum computing. *Reports on Progress in Physics*, 61(2):117–173, 1998.

[7] Fabian Hassler. Majorana qubits, 2014. arXiv: 1404.0897.

[8] Alexei Yu. Kitaev. Unpaired majorana fermions in quantum wires. 44(10S), Oct 2001. arXiv: 0010440.

[9] Anton Akhmerov, Jay Sau, Bernard van Heck, Sebastian Rubbert, and Rafał Skolasiński. `https://topocondmat.org/`. TU Delft 2014-2021, Accessed: 16-05-2021.

[10] August Mikkelsen. Schrödinger-Poisson Calculations for Majorana Devices. Master's thesis, University of Copenhagen, 2017.

[11] T. Ö. Rosdahl, A. Vuik, M. Kjaergaard, and A. R. Akhmerov. Andreev rectifier: A nonlocal conductance signature of topological phase transitions. *Phys. Rev. B*, 97:045421, Jan 2018.

[12] Anna Grivnin, Ella Bor, Moty Heiblum, Yuval Oreg, and Hadas Shtrikman. Concomitant opening of a bulk-gap with an emerging possible Majorana zero mode. *Nature Communications*, 2019.

[13] Yingyi Huang, Jay D. Sau, Tudor D. Stanescu, and S. Das Sarma. Quasiparticle gaps in multiprobe majorana nanowires. *Phys. Rev. B*, 98:224512, Dec 2018.

[14] Ivar Giaever. Energy gap in superconductors measured by electron tunneling. *Phys. Rev. Lett.*, 5:147–148, Aug 1960.

[15] Tudor D. Stanescu and Sankar Das Sarma. Building topological quantum circuits: Majorana nanowire junctions. *Phys. Rev. B*, 97:045410, Jan 2018.

[16] David J. Griffiths. *Introduction to electrodynamics*. Pearson Education Inc., fourth edition, 2015. Edition for Technical University of Denmark and Copenhagen University.

[17] Roland Winkler. *Spin-Orbit Coupling Effects in Two-Dimensional Electron and Hole Systems*. Springer, Berlin, Heidelberg, 2003.

[18] Asbjørn C. C. Drachmann, Rosa E. Diaz, Candice Thomas, Henri J. Suominen, Alexander M. Whiticar, Antonio Fornieri, Sergei Gronin, Tiantian Wang, Geoffrey C. Gardner, Alex R. Hamilton, and et al. Anodic oxidation of epitaxial superconductor-semiconductor hybrids. *Physical Review Materials*, 5, Jan 2021. arXiv: 2009.08190.

[19] Morten Ib Kjærgaard Munk. *MAJORANA BOUND STATES: Readout, decoherence and generalizations*. PhD thesis, 2020.

[20] A. M. Whiticar, A. Fornieri, A. Banerjee, A. C. C. Drachmann, S. Gronin, G. C. Gardner, T. Lindemann, M. J. Manfra, and C. M. Marcus. Zeeman-driven parity transitions in an Andreev quantum dot. pages 1–10, 2021.

[21] Libero J. Bartolotti and Ken Flurchick. *An Introduction to Density Functional Theory*. John Wiley & Sons, Ltd, 1996.

[22] William C Witt, Beatriz G Del Rio, Johannes M Dieterich, and Emily A Carter. Orbital-free density functional theory for materials research. *Journal of Materials Research*, 2018.

[23] Andrew Zangwill. *Modern Electrodynamics*. Cambridge University Press, 2012.

[24] Andrei Postnikov. Lectures on the density functional theory, chapter 1: Thomas-fermi method. `https://www.home.uni-osnabrueck.de/apostnik/`. Accessed: 16-05-2021.

[25] M. P. Hobson K. F. Riley. *Essential mathematical methods for the physical sciences*. Cambridge University Press, third edition, 2016.

[26] B. P. van Zyl, E. Zaremba, and P. Pisarski. Thomas–fermi–von weizsäcker theory for a harmonically trapped, two-dimensional, spin-polarized dipolar fermi gas. *Physical Review A*, 87(4), 2012.

[27] Andrey E. Antipov, Arno Bargerbos, Georg W. Winkler, Bela Bauer, Enrico Rossi, and Roman M. Lutchyn. Effects of gate-induced electric fields on semiconductor majorana nanowires. *Phys. Rev. X*, 8:031041, Aug 2018.

[28] Brandon Peter van Zyl. *Thomas-Fermi-Dirac-von Weizsacker hydrodynamics in low-dimensional electronic systems*. PhD thesis, Queen's University (Canada), January 2001.

[29] Brandon van Zyl, A. Farrell, E. Zaremba, J. Towers, Pawel Pisarski, and D. Hutchinson. Nonlocal kinetic energy functional for an inhomogeneous two-dimensional fermi gas. *Physical Review A*, 89:022503, 02 2014.

[30] Matthias Brack and Rajat K. Bhaduri. *Semiclassical Physics*. Addison-Wesley Publishing Company, Inc., 1997.

[31] W Jones and W H Young. Density functional theory and the von weizsacker method. *Journal of Physics C: Solid State Physics*, 4(11):1322–1330, aug 1971.

[32] COMSOL Multiphysics documentation, version 5.6. `https://doc.comsol.com/5.6/docserver/#!/com.comsol.help.comsol/helpdesk/helpdesk.html`. Accessed: 14-04-2021.

[33] NumPy documentation, version 1.20.0. `https://numpy.org/doc/stable/numpy-user.pdf`. Accessed: 23-04-2021.

[34] FEniCS documentation. `https://fenicsproject.org/documentation/`. Accessed: 23-04-2021.

[35] Anders Logg, Garth Wells, and Kent-Andre Mardal. *Automated solution of differential equations by the finite element method. The FEniCS book*, volume 84. 04 2011.

[36] Luca Bonaventura. Lecture notes: Mathematical methods for engineering: Numerical methods, 2018. MOX - Laboratorio di Matematica Applicata, Dipartimento di Matematica, Politecnico di Milano.

[37] David J. Griffiths. *Introduction to Quantum Mechanics*. Cambridge University Press, second edition, 2017.

[38] J. Robertson. High dielectric constant oxides. *Eur. Phys. J. Appl. Phys.*, Volume 28, Number 3, December 2004.

[39] Semiconductor material database 1. `http://www.ioffe.ru/SVA/NSM/Semicond/index.html`. Accessed: 27-03-2021.

[40] I. Vurgaftman and J. R. Meyer. Band parameters for iii–v compound semiconductors and their alloys. *Journal of Applied Physics 89, 5815*, 2001.

[41] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.B 136, 864*, 1964.

[42] Matthias Ernzerhof and Gustav E. Scuseria. The slowly-varying noninteracting electron gas in terms of its kinetic energy density. J. Chem. Phys. 112, 5270, 2000.

[43] Chun Xiao Liu, Sergej Schuwalow, Yu Liu, Kostas Vilkelis, A. L.R. Manesco, P. Krogstrup, and Michael Wimmer. Electronic properties of InAs/EuS/Al hybrid nanowires. 2020.

[44] S. J. Pauka, J. D.S. Witt, C. N. Allen, B. Harlech-Jones, A. Jouan, G. C. Gardner, S. Gronin, T. Wang, C. Thomas, M. J. Manfra, J. Gukelberger, J. Gamble, D. J. Reilly, and M. C. Cassidy. Repairing the surface of InAs-based topological heterostructures. *Journal of Applied Physics*, 128, 2020.

[45] Georg W. Winkler, Andrey E. Antipov, Bernard Van Heck, Alexey A. Soluyanov, Leonid I. Glazman, Michael Wimmer, and Roman M. Lutchyn. Unified numerical approach to topological semiconductor-superconductor heterostructures. *Physical Review B*, 99(24):245408, 2019.

[46] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan. Band parameters for III-V compound semiconductors and their alloys. *Journal of Applied Physics*, 89, 2001.

[47] Semiconductor material database 2. `https://ecee.colorado.edu/~bart/book/effmass.htm`. Accessed: 28-04-2021.